

A SPECTRO-TEMPORAL ANALYSIS OF SPEECH BASED ON NONLINEAR

Jean Rouat, Sylvain Lemieux and Alain Migneault

Dépt. des sciences appliquées,
Université du Québec à Chicoutimi, 555 Bld de l'université,
CHICOUTIMI, Québec, Canada, G7H 2B1

ABSTRACT

This paper proposes a spectro-temporal analysis based on a bank of cochlea filters in combinaison with a nonlinear operator for amplitude modulation enhancement in the medium and high frequency formants. The output of the spectro-temporal analysis is represented as a 3D image where it is possible to observe very short-term speech transitions and formant modulations. With such analysis, it is possible to obtain patterns characteristics of phonemes and transitions between phonemes, which can not be obtained by using other speech analysis (FFT, LPC) techniques. The paper presents 3D images of vowels, where the amplitude modulation of the formants is clearly visable.

1.INTRODUCTION

Research in speech analysis is recognized to be an important field 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 proposes an analysis to enhance the modulation properties in speech.

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 understanding speech and / or in the design of an efficient speech analysis.

2.MODULATED TONE PERCEPTION

Since the auditory system does not resolve the high frequency components, the temporal features of vowel-like sounds are coded similarly as those of amplitude-modulated tones. Furthermore, research work on automatic demodulation of speech can be motivated by the hypothesis proposing that the human brain has neural cells which specialize in Amplitude Modulation (AM) and Frequency Modulation (FM) detection [3][18][19]. More recently, Schreiner and Langner [6][17] have studied the representation of amplitude modulation in the inferior colliculus of cats and have shown that the inferior colliculus of the cat contains a highly systematic topographic representation of amplitude modulation paremeters.

3. BASILAR MEMBRANE NONLINEARITIES

Nonlinearity and perception of intermodulation distortion products (f_1-f_2 , $2f_1-f_2$, etc.) are a pressing issue in hearing research and it is not easy to understand exactly the origin of these nonlinearities. Recently Robles, Ruggero and Rich have observed distortion products on chinchilla basilar membrane by using a laser-velocimetry technique [16]. 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 SPEECH PROCESSING

Non linear processing

The proposed analysis attempts to consider the automatic demodulation of the signal, before it is transformed in neural pulses in the cochlea. In fact, we will show that nonlinear operations of the signal create distortion products and can enhance the modulation properties of the signal. Two nonlinear operators will be included in the proposed analysis to enhance the Amplitude Modulation observed with vowels.

Nonlinear filtering seems to be very attractive and much work has been done in that field, refer to [9] for examples. More recently, one can cite the work by P. Maragos et al [7] where it is shown that the nonlinear operator, called Teager energy operator [5], allows AM and FM demodulation. Furthermore, L. Atlas and J. Fang [1] have shown that quadratic detectors allow for a better representation of speech in the context of a noisy pitch tracker.

The originality of the present work resides in the combination of a perceptive bank of filters with nonlinear operators to obtain a 3D representation of speech with the AM information enhanced.

Nonlinear operators

J.F. Kaiser [5] proposes the Teager energy operator as being able to extract the energy of a signal based on mechanical and physical considerations. It has been shown [7] that this operator is able to track either the amplitude of an A.M. signal or the frequency of an FM signal.

Another nonlinear operator has been proposed [14] to take into consideration the changes in the instantaneous signal power in the cochlea. This operator, called "Dyn", shows the ability to enhance the AM-FM modulation in speech.

Generally speaking, nonlinear operators are simple tools with the ability to modify the signal spectrum by combining the spectrum information. This ability is particularly interesting for AM or FM demodulation and for spectrum shifting, which are not easy to perform with standard linear techniques.

Figure 1 illustrates the output of the Teager Energy and Dyn operators for two tones. The first section of figure 1 is a 600Hz tone, the second section is a 1000Hz tone. Section 3 is the sum of the 600Hz and 1000Hz tones. Sections 4 and 5 are respectively the output of the Dyn and Teager energy operators for the signal from section 3.

Let us consider the combination tone defined as :

$$s(t) = A_1 \cos(\omega_1 t) + A_2 \cos(\omega_2 t).$$

By using the analog version of the Teager energy operator [4], one can show that:

$$\begin{aligned} \text{Teager}[s(t)] &= (A_1 \omega_1)^2 + (A_2 \omega_2)^2 \\ &+ (A_1 A_2) \left(\frac{\omega_1^2 \omega_2 + \omega_1 \omega_2^2}{2} - \omega_1 \omega_2 \right) \cdot \cos[(\omega_1 + \omega_2) t] \\ &+ (A_1 A_2) \left(\frac{\omega_1^2 \omega_2 + \omega_1 \omega_2^2}{2} + \omega_1 \omega_2 \right) \cdot \cos[(\omega_1 - \omega_2) t] \end{aligned} \quad (1)$$

The amplitude difference between the two tones from the Teager output is equal to $2A_1 A_2 \omega_1 \omega_2$. Therefore, the $\omega_1 - \omega_2$ component will largely dominate in comparison with the $\omega_1 + \omega_2$ component, as it is observed in section 5 from figure 1 where $\omega_1 - \omega_2 = 2\pi(1000-600)$ rad/s.

Similarly, by using the analog form of the Dyn operator [13], one can show that :

$$\begin{aligned} \text{Dyn}[s(t)] &= -\frac{A_1^2 \omega_1}{2} \sin(2\omega_1 t) - \frac{A_2^2 \omega_2}{2} \sin(2\omega_2 t) \\ &- A_1 A_2 \left(\frac{\omega_1 \omega_2 + \omega_1^2}{2} \right) \sin[(\omega_1 + \omega_2)t] \\ &- A_1 A_2 \left(\frac{\omega_1 \omega_2 - \omega_1^2}{2} \right) \sin[(\omega_1 - \omega_2)t] \end{aligned} \quad (2)$$

By comparing equation (2) with figure 1, we observe that the component $\omega_1 + \omega_2 = 2\pi(1000+600)$ rad/s is predominant in the output of Dyn for the composite signal.

In summary, nonlinear operators are simple and powerful tools to obtain distortion components from a sum of pure tones and might be used in speech processing where, some of the distortion components might be perceptively important.

5. THE ANALYSIS OF SPEECH

In this section, we will describe how a perceptive filterbank has been used in conjunction with the Teager energy or Dyn operators to generate a 3D representation of speech where the amplitude modulation of formants has been enhanced.

Filtering

The actual version of the analyzer is comprised of a bank of twenty-four filters centred on 330Hz to 4700Hz. These filters partially simulate the frequency analysis performed by the cochlea. These are rounded exponential filters with the Equivalent Rectangular Bandwidths (ERB) proposed by Patterson [11] and Moore and Glasberg [10]. 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 [2] the output signal $s_i(t)$ from channel i can be considered to have been modulated in amplitude and phase with a carrier frequency of f_i .

$$s_i(t) = A_i(t) \cos [\omega_i t + \phi_i(t)] \quad (3)$$

$A_i(t)$ is the modulating amplitude and $\phi_i(t)$ is the modulating phase. It should be noticed that equation (3) is true only for a bandpass signal (bandwidth of $A_i(t)$ and $\phi_i(t)$ small in comparison to f_i).

AM - FM demodulation

The output of the analog version of the Teager energy operator [5] to $s_i(t)$ is related to the modulating amplitude multiplied by the instantaneous frequency by the term $[A_i(t) (\omega_i +$

Error!

Similarly, the output of the low-pass filtered Dyn operator [13] to $s_i(t)$ is equal to the fluctuations of $A_i^2(t)$, $(\frac{1}{4} \frac{d}{dt} [A_i^2(t)])$.

The speech data

The speech has been sampled to 32 kHz with a 16 bit A/D converter in a computer room. When needed, a white Gaussian noise has been added to the speech with a SNR of 10 dB. One male and one female have pronounced the cardinal vowels /a/, /i/ and /u/ in the CV context, with the consonants being /b/, /d/, /g/, /p/, /t/, /k/.

Experiment with the Teager Energy operator

After filtering, the output of each channel has been processed by the Teager energy operator and the square root of the raw data is plotted as three dimensional images. No smoothing or low-pass-filtering of Teager energy output has been made. As a matter of fact, $A_i^2(t) [\omega_i +$

Error!

Experiment with the Dyn operator

The output of each channel has been processed by the Dyn operator, then Dyn has been low-pass-filtered to obtain the expression $\frac{1}{4} \frac{d}{dt} [A_i^2(t)]$ before plotting the 3D representation.

Analysis of the output

Fig. 2, and fig. 3 present the output of the 3D analysis of a /a/ taken from the french syllable "da" for the Teager energy operator and the 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 low-pass-filtered Dyn.

Both operators give a rich representation of the /a/ with a typical modulation pattern in the formant. Teager energy operator enhances the high frequency channels in comparison to Dyn. Both Teager and Dyn operators allow an easy detection of the AM modulation in the formant.

Figures 4 and 5 are the 3D representation of a /i/ taken from the syllable "di". Figure 4 is the pronunciation by a male speaker and figure 5 by a female speaker. The second and third formants of the /i/ are clearly modulated in amplitude for both speakers. The frequency of the AM modulation in the third formant is lower than in the second formant. The second and third formants energy, in comparison to the first formant, is more important for the female speaker than for the male speaker. The first formant is not modulated in amplitude for both speakers.

7. CONCLUSION

We have proposed a new speech analysis based on nonlinear operators for AM "demodulation". We have presented results with the Teager energy and the Dyn operators. With such analysis it seems possible to obtain patterns characteristics of phonemes and transition between phonemes, which can not be obtained by using other speech analysis (FFT, LPC) techniques.

The presented figures show clearly modulation patterns in the /a/ and /i/ vowel. Such analysis can be used as a new tool by the speech community to observe and characterize speech in terms of Amplitude Modulation.

The filterbank is comprised of twenty-four filters, this is a minimum number. A bank of sixty-four filters would give better results.

In conclusion, the proposed analysis is fully parallel and can be implemented in real time on a parallel VLSI. Other experiments are in progress on automatic demodulation of noisy speech.

Acknowledgment:

This work has been supported by the NSERC of Canada under Grant N° OGP0042386, by the Canadian Microelectronic Corporation, 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 Y.C. Liu.

8. References

[1] L. Atlas and J. Fang. "Advantages of General Quadratic Detectors for Speech Representations." in "Visual Representations of Speech Analysis.", edited by M. Cook and S. Beet, J. Wiley, 1992.

- [2] A.B. Carlson "Communication systems: An Introduction to Signals and Noise in Electrical Communication". Mc Graw Hill, 1986.
- [3] R.B. Gardner and J.P. Wilson. "Evidence for direction-specific channels in the processing of frequency modulation". J. Acoust. soc. Amer. 66, 704-709, 1979.
- [4] J.F. Kaiser. " On Teager's energy algorithm and its generalization to continuous signals." in proc. IEEE DSP workshop, New Paltz, NY, september 1991.
- [5] J.F. Kaiser. "On a simple algorithm to calculate the 'energy' of a signal". Proceedings of IEEE-ICASSP'90, Albuquerque, 381-384.
- [6] G. Langner and C.E. Schreiner . "Periodicity coding in the inferior colliculus of the cat. Neuronal mechanisms". J. Neurophysiol. Vol. 60, no 6, 1799-1822, 1988.
- [7] P. Maragos, T. Quatieri and J.F. Kaiser. "Speech nonlinearities, modulations and energy operators". Proceedings of the IEEE - ICASSP'91, Toronto, 421-424.
- [8] P. Maragos, J.F. Kaiser and T.F. Quatieri. "On Amplitude and Frequency Demodulation Using Energy Operators". Accepted for publication in IEEE SP, April 1992.
- [9] V.J. Mathews. " Adaptive Polynomial Filters". IEEE SP magazine, vol 8, nb 3, 10-26, 1991.
- [10] B.C.J. Moore and B.R. Glasberg. "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns." J. Acoust. Soc. Am. 74, 750-753, 1983.
- [11] R.D. Patterson . "Auditory filter shapes derived with noise stimuli." Jour. Acoust Soc. Amer., 59 , 3, 640 - 654, 1976.
- [12] R. Plomp . "Effect of amplitude compression in hearing aids in the light of the modulation transfer function." J. Acous. Soc. Amer. 83 (6), 2322-2327, 1988.
- [13] J. Rouat. "Nonlinear operators for speech analysis". in "Visual Representations of Speech Analysis.", edited by M. Cook and S. Beet, J. Wiley, 1992.
- [14] J. Rouat. "Dyn: a nonlinear operator for speech analysis." Université du Québec à Chicoutimi, dépt des sciences appliquées, internal report, 1991.
- [15] J. Rouat, Y.C. Liu and S. Lemieux. " A nonlinear analysis for clean and noisy speech". Canadian Acoustics, vol. 19, nb 4, sept. 1991.
- [16] M.A. Ruggero and N.C. Rich. "Application of a commercially-manufactured Doppler-shift laser velocimeter to the measurement of basilar-membrane vibration" Hear. Res. 51, 215-230, 1991.
- [17] C.E. Schreiner and G. Langner ."Periodicity coding in the inferior colliculus of the cat. Topographical organization". J. Neurophysiol., Vol 60, no 6, 1823-1840, 1988.
- [18] B.W. Tansley and J.B. Suffield. "Time course of adaptation and recovery of channels selectively sensitive to frequency and amplitude modulation." J. Ac. Soc. Am. '74, 765-775, 1983.
- [19] G.H. Wakefield and N.F. Viemeister. "Selective adaptation to linear frequency-modulated sweeps: Evidence for direction-specific FM channels?" J. Ac. Soc. Amer. 75, 1588-1592, 1984.