LATERAL INHIBITION AND SPEECH SIGNAL PROCESSING

DANG* V.C., CARRE R.

Laboratoire de la Communication Parlée, ICP Unité Associée au CNRS,
INPG-ENSERG, 46 Avenue Félix Viallet,
38031, GRENOBLE CEDEX, FRANCE.
*Present address : Khoa vô tuyên, DHBK, HANOI, VIETNAM.

ABSTRACT :

Lateral inhibition is a side-band effect of excitation of the auditory system by a complex signal. Indeed, single neuron response is modified by the signals issued by surrounding neurons due to the complex stimulation. In this paper, we present works on this subject using a simplified model over natural and synthetic speech sounds. A spectral lateral inhibition is used to enhance spectral peaks. Preliminary tests on temporal lateral inhibition (lateral inhibition in time-domain) show an enhancement of time-domain contrasts. This information might be used to find stable regions in the speech signal.

1. INTRODUCTION

In the past years, it has been recognized the existence of a lateral inhibition function in neuronal processings and several works have been developed on the modelling of this function (GREENWOOD & MARUYAMA - 1965, GREENWOOD & GOLDBERG - 1970, MORISHITA & al. - 1972, TOKURA & al. - 1977, CAELEN - 1979, VOIGHT & YOUNG -1980, PALMER & EVANS - 1982, MARTIN & DICKSON - 1983, SHAMMA - 1985). In short, single neuron response is modified by the signals issued by surrounding neurons due to the complex stimulation.

KARNICKAYA & al. (1973) have applied a "lateral inhibition" model on the auditory spectrum equivalent and have observed that spectral contrasts are increased. They have used a three-range window : a central positive one and two lateral negative ones, gliding in the frequency domain.

MORISHITA & al. (1977), SHAMMA (1985) have tested neuron network models.

LEBEDEV & al. (1985) have built a performant recognition system by taking into account the time-domain and frequency-domain masking effects.

In this work, a simplified inhibition model similar to that of Karnickaya's is tested, in the frequency domain and in the time domain, to point out contrast effects on the spectrum. This three-range window model can be compared with the cepstral technique where the inverse FFT + square windowing + direct FFT block corresponds to a sin(x)/x operator with a positive central lobe and two main negative lobes.

2. EXPERIMENTS

Original speech signal is low-pass filtered at 5 KHz, sampled at 10 KHz, weighted by a Hamming window and processed via FFT. The spectral components e(t,j), at the FFT output(t = time, j = frequency) are then processed by a lateral inhibition system.

In the frequency domain, the spectral lateral inhibition filter contains one central region and two lateral inhibition regions, as represented in figure 1.

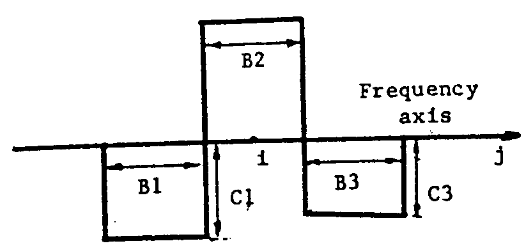

Figure 1. Spectral lateral inhibition filter.

The filter output S(t,i) is the weighted sum of the inputs e(t,j) in the central region minus the sums of the two lateral inhibition regions :

$$S(t,i) = -C1\sum_{j\in B1} e(t,j) + \sum_{j\in B2} e(t,j) - C3\sum_{j\in B3} e(t,j)$$

C1, C3 are amplitude constants.

The three range filter is applied on the FFT spectral components and the filter is gliding on the frequency scale. B1, B2 and B3 are set in a Bark scale.

In the time domain, two filters have been tested :

Type 1 : the time domain filter is exactly corresponding to the frequency domain filter described above :

$$S(t,i)' = -D1\sum_{t\in T1} e(t,j) + \sum_{t\in T2} e(t,j) - D3\sum_{t\in T3} e(t,j)$$

D1, D3 are amplitude constants and the T constants are duration constants. Figure 2 represents the time-domain lateral inhibition law, which is a modification of Lebedev's time masking curve (LEBEDEV & al. (1985)).

Type 2 : the output element $S_2(i)$ in a time-domain lateral inhibition filter is computed over the sum of the absolute values of differences between the spectral components processed at time i and i-1 (the output of the first element $S(i,j)$). Equations a and b define $S_2(i)$

a) $$S_1(i) = \sum_{j=1}^{N} |S(i,j) - S(i-1,j)| \;;$$

b) $$S_2(i) = -D1\sum_{j\in T1} S_1(j) + \sum_{j\in T2} S_1(j) - D3\sum_{j\in T3} S_1(j)$$

i, j, T1, D1, T2, T3, D3 are defined on figure 2.

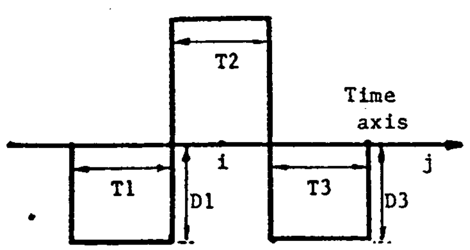

Figure 2. Time-domain lateral inhibition law.

3. RESULTS AND DISCUSSIONS

Spectral lateral inhibition.

a) Study of the model parameters.

In order to study the role of the model parameters. /a/, /i/, /u/ vowel spectra were calculated for different values of one parameter among the others. The objective was to find the right value corresponding to a better contrast effect on the spectrum. The optimal values for the window ranges are 1 Bark, and the amplitudes C1, C2, C3 are -0.7, 1, -0.3. These values are closed to those proposed by KARNICKAYA.

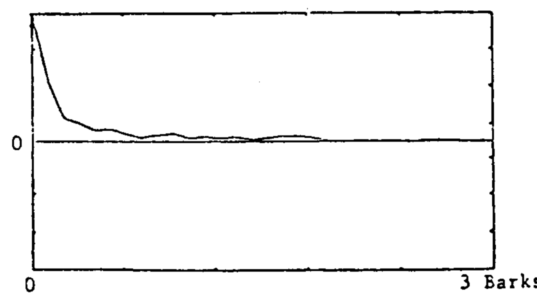

Figure 3. Evolution of the spectral distance (vowel /a/) for different B1 values.

Such parameter values were tested to verify the good stability of synthetic vowel spectrum representation. The euclidian distance between two successive spectra was calculated for different values of each parameter (figure 3). This distance has a first minimum when B values equal 1 Bark and when C values are around -0.7, 1, -0.3.

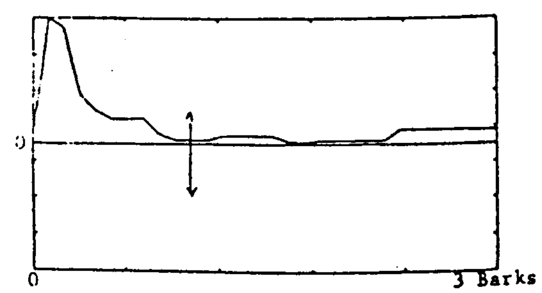

Figure 4. Distance between the formant values (for synthetic vowels) and the spectral peak values, for different B1 values.

The parameters were also tested to verify the good acuracy of the spectral peaks. For different synthetic vowels with specified formant frequencies, the distances between these frequencies and the peak frequencies of the spectral representation were calculated. Again the distances have a first minimum at around 1 Bark for the B values and at -0.7, 1, -0.3 for the C values (figure 4).

b) Results on synthetic vowel signals.

Figure 5 shows the spectral representation obtained by FFT (curve 1), FFT + 1 Bark integration (curve 2), cepstral technique (curve 3) and FFT + lateral inhibition.

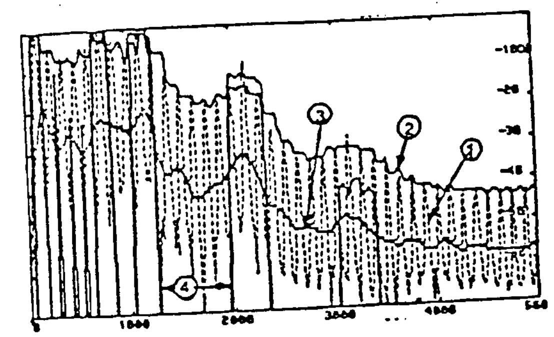

Figure 5. Spectral lateral inhibition for /a/ vowel (curve 4).

The parameters of the lateral inhibition model are : 1 Bark for the B values and -0.3, 1, -0.7 for the C values. The spectral contrast is clearly increased.

In the case of noisy vowels, spectral peaks are better represented in the case of lateral inhibition processing (Figure 6).

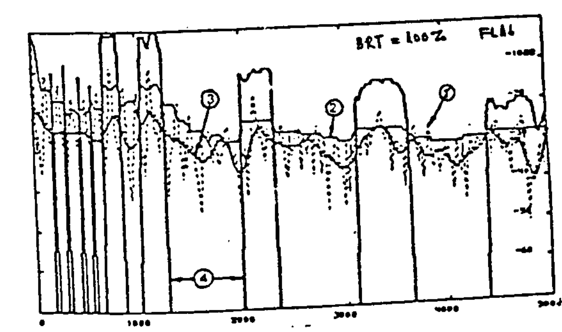

Figure 6. Spectral lateral inhibition for noisy /a/ vowel (noise level 100%) - curve 4 . Curve 1 : FFT, curve 2 : FFT + 1 Bark integration, Curve 3 : cepstral representation.

c) Results on natural CVCVC sounds.

The use of the spectral lateral inhibition clearly increases the spectral contrast (figure 7). When the central range of the model is only one FFT value, the contrast is more important and the harmonic structure appears mainly for low frequencies (figure 8).

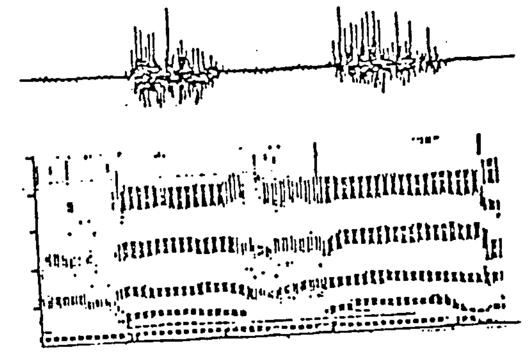

Figure 7. FFT + spectral lateral inhibition (CVCVC : /babab/).

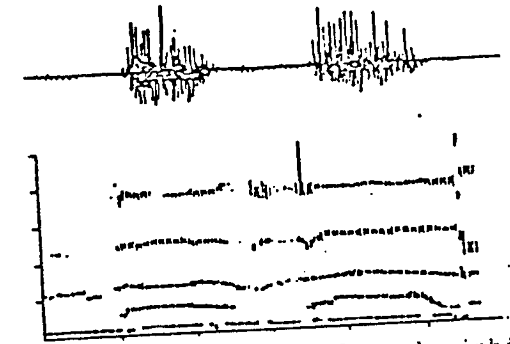

Figure 8. FFT + reinforced inhibition (CVCVC : /babab/)

Temporal lateral inhibition.

The type 2 representation is given on figure 9 with duration ranges of 5 ms. Lateral inhibition gives peaks at the place of temporal transitions. This system could be used for event detection.

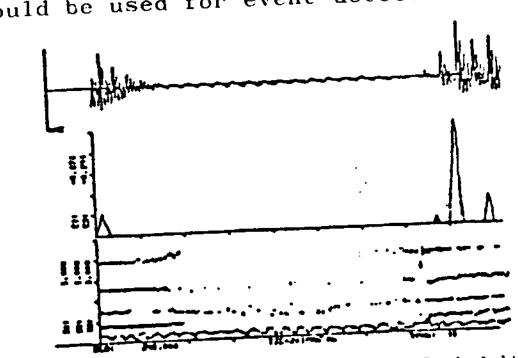

Figure 9. Time domain lateral inhibition. Temporal range = 5 ms, CVCVC : /babab/, /aba/ part.

4. CONCLUSIONS

The results obtained show that lateral inhibition is able to increase temporal and spectral irregularities. Increased spectral irregularities enhance the spectral peaks. Thus, the speech spectrum is simplified. According to the parameter values of the model, the low frequency harmonic structure can be observed.

In the time domain, according to the parameter values of the model, lateral inhibition enhances either the boundaries of the stationary sounds or small temporal events.

5. REFERENCES

(1) Chistovich L.A, Lublinskaya V.V., Malinnikova T.G., Ororodnikova E.A., Stoljarova E.I and Zhukov S. Ja. (1982).
Temporal processing of peripheral auditory patterns of speech.
In "Representation of speech in the peripheral auditory system", ed. Carlson R. and Granstrom B., Amsterdam, 165-181.

(2) Dang V.C., Carré R. and Tuffelli D. (1986).
Research on preprocessing by a lateral inhibition.
12th. Int. Cong. on Acoustics, Montreal.

(3) Karnickaya E.G, Mushnikov V.N., Slepokurova N.A. (1973).
Auditory processing of steady-state vowels.
Symp. on Auditory Analysis and Perception of speech, Leningrad.

(4) Lebedev V.G. and Zagoruiko N.G. (1985).
Auditory perception and speech recognition.
Speech communication, **4**, 97-103.

(5) Morishita I. and Yajima A. (1972).
Analysis and simulation of networks of mutually inhibiting neurons.
Kybernetic, **11**, 154-165.

(6) Tokura T. and Morishita I. (1977).
Analysis and simulation of double-layer neural network with mutually inhibiting interconnections.
Biol. Cybernetics, **25**, 83-92.

(7) Shamma S.A. (1985).
Speech processing in the auditory system, II: Lateral inhibition and the central processing of speech evoked activity in the auditory nerve.
J. of Acoust. Soc. of Am., **78**, 1622-1632.