POLYTONAL ANALYSIS-SYNTHESIS SYSTEM FOR SPEECH ENHANCEMENT

V. MAKHONIN

Institute for problems of information transmission
USSR Academy of Sciences
101447 ГСП-4 Moscow USSR

ABSTRACT

Most of techniques for speech enhancement using noise suppression suppress discontinuously modulated speech oscil - lation too.At result of discontinuously modulated oscillations suppression the quality of selected speech signal is decreased.One way of enhancing speech in an additive noise is to perform a functional decomposition of a frame of noisy speech and to attenuate a particular transformed component depending on how much the measured component pulsation power exceed an estimate of the background noise.Using a Walsh-Hadamard blocks connected by strings of zeroes results in a new class of suppression curves which permits a tradeoff of noise suppresion against speech distortion.The algorithm has been implemented in "Eclips C-330" minicomputer.

INTRODUCTION

The security of vocoders,speech recognition devises and speech synthesizers is less than natural one . It was found by speech testing that technicians ignore high frequency part of telephone spectra / 1 /. The level of high frequency oscillations is low,but these oscillations transmit an important information for human hearing becouse of voise modulations.

The signal preemphasis before an A/D convertion and the discontinuous demodulation of speech oscillations select speech data for proper display and hearing expe - riments.

Different techniques have been compared for enhancing the noisy speech.The results seem to point out the superiority of block-cascade Walsh-Hadamard transformations especially concerning heavy noise environment. Lines of Hadamard matrix are strings and elements of these strings are +1 and -1. Some of these lines seem like clipped harmonics , while others seem like clipped phase shift keyed modulated oscillations.

Thus one string of ± 1 had been taken from Hadamard matrix repeated periodically together with string of zeroes produce sequence of block clipped waves. The set of such sequenses and a frame of samples have been taken from speech signal produce the set of scalar products.

Next step in frame processing is neglecting of oscillation components those pulsations power are less then corresponding background thresholds. Neglecting reduces noise components , while strongest poly - tonal components represent voised speech signal and its waveform microvariations by sequence of such accords.

The work presented here is continuation of works have been carried out ten years ago during a stage in ULB / 2 / and later / 3,4 /. Concerning speech signal microvariations representation few researches have been done / 5,6 /. One was performed by M.Rohtla et al. / 7 /.

OUTLINE OF THE POLYTONAL TRANSFORMATIONS

The described computational tehnique has been employed according to the frame by frame processing mode. While samling frequency = 20 kHz and frame duration = 51.2 msec each frame consist of 1024 samples.Frames are overlapped on half of frame size,i.e. 25.6 msec.

The frame is transformed into set of scalar products of speech signal samples and elements of an elementary subsequence. While transformation is block-cascade Walsh-Hadamard transformation , the elementary subsequence consist of $\pm$ 1 inside and 0-es outside of blocks , accordingly.

For example , if tone period = 128 samples , line seven from an Hadamard transform matrix formed of eight lines and eight columns has been chosen and the chosen block is second , the elementary subsequence is represented as follows:

0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,-1,-1,+1,+1,-1,-1,+1,+1,+1,+1,-1,-1,+1,+1,-1,-1, 0,0,0,(in all 128-16=112 zeroes),-1,-1,+1,+1,-1,-1,+1,+1,+1,+1,-1,-1,+1,+1,-1,-1, 0,0,0,(in all 112 zeroes again), -1,-1, et cetera up to 1024th.

After transformations of 1024 samples into 1280 scalar products (20 tones*8 lines * 8 blocks = 1280 elementary subsequences) have been fulfilled the second cascade of transformations to be fulfilled. By those second transformations one get a set of estimations pitch pulsations sincronouosly to tones scanned. Estimations taken from this set are compared with corresponding thresholds , those are greather then thresholds to be seleated,others to be neglected and their values be - come equal zeroes.

Next steps are related with choises of tones from scanned scale. To choose tones estimations corresponded to scanned tone collected together to transform in logarithmic scale and to compare those partial pulsation levels with corresponding thresholds and select tones those levels are greather... and so on.

It is possible to print the results of computations in format as data represented in Tables N° N° 1,2,3,4,5,6 or to rewrite a linear combinaton of selected oscillations on the disk memory and to convert records by D/A converter for hearing experiments.

So ,a usage of described computation process permits to select so many tones as nessesary to represent signal microvaritions and to enlarge frame size to suppres noise oscillations.

COMMENTS TO TABLES

A table of results of polytonal analysis consists of two parts,left with "stars" instead of zeroes and integer nombers I2, which represent pulsation level distribution between lines of Hadamard matrix. Signal time is growing from row to row on the amount of overlap between frames,i.e. 25.6 msec.

SUMMARY

A polytonal analysis-synthesis system has been described , which has appliation to robust speech processing. The experi - ments using the new model of speech sig - nal indicate its power in synthesizing natural sounding voised signals.

REFERENCES

/1/ Речевые тесты и их применение. Изд-во МГУ. 1986г.,стр.64

/2/ Makhonine V. On the representation of discontinuous speech acoustical events. Rapp.d'activ.de l'inst.de phon.RA12/1 Brux.1978.

/3/ Махонин В.А. Изучение микровариации речевого сигнала.Франко-Советский симпозиум по исследованию речевой информатики. Гренобль I98Iг. стр.303-3I2

/4/Д.Отесер и др. Экспериментальная методика наблюдения микровариаций.Советско-Французский симпозиум"Акустический диалог человека с машиной.М.I984г.стр.III-II6

/5/Пирогов А.А.Предисловие к переводу . ТИИЭР,Том 73,№II ,М.,"Мир" I986г.,стр.3

/6/Арнольд В.И. Теория катастроф. Изд-во МГУ ,I983г.,стр.24-25

/7/ Рохтла М.,Раудсепп М. Зависимость качества синтезированной речи от тонкой структуры изменения основного тона.Тезисы АРСО-I4.ЧастьI.Изд-во КПИ,Каунас I986г. стр.57-58.

/8/ H.Nagabuchi,T.Kobayashi,F.Itakura Effect of the comb-filtering noise reduction method in speech analysis-synthesis processing in noisy environments. Rep. Acoust.Soc.Jap.1980. S80-54.

APPENDIX

Table 1
File: "HA" female speaker
Length of minimal pitch period=77
Step of pitch period increasing=1
Threshold for modul.selection=0.25

*************9*****	95	2	0	0	0	0	0	0
9******************	94	2	0	0	0	0	0	0
92*****6***********	95	2	o	0	0	0	0	0
943*723*31**1	89	6	0	1	0	0	0	0
91*74**51******	88	5	0	1	0	0	0	0
9*338**5*******	88	6	0	1	0	0	0	0
938**********	88	5	0	1	0	0	0	0
92*6*351*****	90	5	0	1	0	0	0	0
********9*74*******	81	12	1	2	0	0	0	0
*******9***********	87	7	2	0	0	0	0	0
*********96********	86	8	1	1	0	0	0	0
**********9********	82	12	0	2	0	0	0	0
**********9********	88	7	1	1	0	0	0	0
**********9********	86	8	1	1	0	0	0	0
*********9246******	33	50	5	5	0	0	1	1
*********94**4*****	61	25	3	6	0	0	0	1
******92***********	86	7	3	0	0	0	0	0
****94*52**********	76	15	2	2	0	0	0	0
917************	73	20	0	3	0	0	0	0
92*************	88	7	0	1	0	0	0	0

Table 2
File: "MA" female speaker
Length of minimal pitch period=77
Step of pitch period increasing=1
Threshold for modul.selection=0.25

92*55********	94	3	0	0	0	0	0	0
*9*****************	89	5	0	0	0	0	0	0
********97*3********	87	6	0	2	0	0	0	0
********95*3********	90	5	0	1	0	0	0	0
********966*********	91	5	0	1	0	0	0	0
*********91*********	85	9	0	2	0	0	0	0
********966*********	78	14	1	3	0	0	0	0
*********91*********	76	15	1	3	0	0	0	0
*********92*********	90	6	0	1	0	0	0	0
*********97*********	85	5	4	1	0	0	1	0

```
**********9**********   76 16 1 3 0 0 0 0
**********92*********   87  7 2 1 0 0 0 0
***********91********   88  7 0 1 0 0 0 0
**********9**********   89  5 2 0 0 0 0 0
*********945*********   54 32 4 4 0 0 1 1
******9**************   88  6 2 1 0 0 0 0
**98*****************   80 12 3 2 0 0 0 0
9********************   85  9 1 1 0 0 0 0
9********************   78 15 1 2 0 0 0 0
```

Table 3
File: "MA" male speaker
Length of minimal pitch period=110
Step of pitch period increasing=2
Treshold for modul.selections=0.1

```
****943*************   89 66 0 1 0 0 0 0
******915***********   91  5 0 1 0 0 0 0
*****9**************   82 11 0 1 0 0 0 0
****91**********3***   74 13 4 1 0 1 0 1
****9***************   70 13 5 2 0 0 5 0
**93*3*7************   44 3111 2 1 1 5 1
**988***************   44 2314 4 1 7 2 0
****91*4************   45 1910 9 1 4 5 1
*****978************   51 1323 1 1 2 4 0
*******98***********   68  9 7 2 1 3 5 0
**********96*5******   56 20 9 2 0 1 6 0
***************94327   61 17 4 6 1 4 1 1
```

Table 4
File "МИ " male speaker
Length of minimal pitch period=112
Step of pitch period increasing=2
Treshold for modul.select.=0.1

```
*****94*************   92 4 0 0 0 0 0 0
***********91*******   88 7 0 1 0 0 0 0
****9***33241*******   90 6 0 1 0 0 0 0
*******9552*********   89 6 0 1 0 0 0 0
*****967************   91 5 0 1 0 0 0 0
****9283******5*****   85 8 0 2 0 0 0 0
**9*87**************   83 7 0 2 0 1 1 0
****97**************   8310 0 2 0 0 0 0
****98**************   85 7 0 1 0 1 0 0
*****97*************   8712 0 2 0 0 0 0
*****9452***********   8211 0 2 0 0 1 0
******96**5*********   8211 0 2 0 0 0 0
***********944******   86 8 0 2 0 0 0 0
****************9*4*   87 8 0 1 0 0 0 0
```

Table 5
File " НАДУВАТЬ ",male speaker,heavy noise environment.
Length of minimal pitch period=166
Step of pitch period increasing=3
Treshold for modul.select.=0.25

```
9***5***************   60 13 9 8 2 2 0 0
***97***************   60  7 5 0 4 4 4 0
*****96*************   36 33 511 7 2 1 1
**7*4*********5*****   52 25 8 3 2 2 2 1
*94*8***************   46 31 5 7 2 2 2 1
9**3****************   74  9 3 5 3 0 0 0
9862*51*1***********   64 15 4 4 2 2 2 2
96*7************2***   55 18 8 6 4 1 2 2
92****83************   58 21 8 4 2 1 1 0
976****************2   61 19 8 4 1 0 1 2
98*4****************   61 21 4 4 2 1 2 1
96****6*************   55 15 5 5 4 6 2 2
*********91*********   64 14 9 3 1 1 2 1
```

```
98**4********4******   38 27 17 5 2 3 3 1
957*2***************   40 33 11 3 1 2 3 1
92******************   38 27 21 2 0 1 6 0
9******3************   12 53 10 8 5 2 2 2
9*********8*3*******   23 28 18 126 5 3 1
```

Table 6
File "ПЛЯСОВОЙ ",male speaker
Length of minimal pitch period=170
Step of pitch period increasing=3
Threshold for modul.select.=0.15

```
***9****************   90  1   1 1 1 0 0 1
98******************   94  1   0 0 0 0 0 0
9*******************   60 17   510 1 2 1 1
95*******3**********   57 25   7 2 1 2 1 0
9438*2**************   72 16   3 2 1 0 0 0
9***5****2**********   61 22   2 7 0 0 0 1
9*******************   54 30   2 6 0 1 0 1
984****22***********   69 17   3 4 1 1 1 0
********************    0  0   0 0 0 0 0 0
********************    0  0   0 0 0 0 0 0
********************    0  0   0 0 0 0 0 0
9*******************   42 12 1323 0 1 1 3
9*****************7*   24 40 19 5 2 2 2 1
9*******************   53 33   5 3 0 0 1 0
9*******************   22 44 18 7 0 0 3 1
9******7************   33 43   9 6 501 0 1
96*****3****2****2**   51 31   6 5 0 1 1 1
*954*2**********3**4   56 26   8 3 0 1 1 0
9****1***********7**   15 41 30 2 0 2 4 0
*9******************   6  51   3 2 0 3 5 0
***9******54********   14 63   9 3 0 5 1 1
*****9**************   21 14 48 2 0 3 7 0
**********9*********   55 23   9 4 1 0 2 0
**************9467**   51 31   7 5 0 0 1 1
***************9****   46 39   3 7 0 0 0 1
```