PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
G10L 9/16

A1

(11) International Publication Number: WO 96/18996

(43) International Publication Date: 20 June 1996 (20.06.96)

(21) International Application Number: PCT/GB95/02943

(22) International Filing Date: 15 December 1995 (15.12.95)

(30) Priority Data:
94309391.4 15 December 1994 (15.12.94) EP
(34) Countries for which the regional or international application was filed: AT et al.

(71) Applicant *(for all designated States except US)*: BRITISH TELECOMMUNICATIONS PUBLIC LIMITED COMPANY [GB/GB]; 81 Newgate Street, London, EC1A 7AJ (GB).

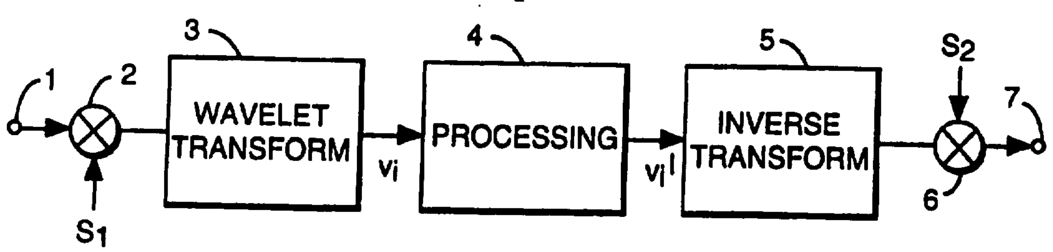
(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: SUMMERFIELD, Stephen [GB/GB]; Dept. of Engineering, University of Warwick, Coventry CV4 7AL (GB).

(74) Agent: LLOYD, Barry, George, William; BT Group Legal Services, Intellectual Property Dept., 13th floor, 151 Gower Street, London WC1E 6BA (GB).

(81) Designated States: AL, AM, AT, AU, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TT, UA, UG, US, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG), ARIPO patent (KE, LS, MW, SD, SZ, UG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: SPEECH PROCESSING



(57) Abstract

A clipped input speech waveform is divided (3) into a plurality of a series of signals by means of a wavelet transform such as the Daubechies wavelet transform, which are then scaled or otherwise processed (4) to reduce the effects of clipping, prior to reconstruction (5) of the speech waveform using the inverse transform.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgystan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LU	Luxembourg	TD	Chad
CS	Czechoslovakia	LV	Latvia	TG	Togo
CZ	Czech Republic	MC	Monaco	TJ	Tajikistan
DE	Germany	MD	Republic of Moldova	TT	Trinidad and Tobago
DK	Denmark	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	US	United States of America
FI	Finland	MN	Mongolia	UZ	Uzbekistan
FR	France			VN	Viet Nam
GA	Gabon				

SPEECH PROCESSING

The present invention is concerned with processing of speech signals, particularly those which have been distorted by amplitude-limiting processes such as clipping.

Apart from its obvious effect on perceived speech quality, clipping in a telecommunications system is disadvantageous in that it reduces the dynamic range of the signal which can adversely effect the operation of echo cancellers.

According to the present invention there is provided an apparatus for processing speech comprising:

means to apply to a speech signal a wavelet transform to generate a plurality of transformed components;

means to modify the component such as to increase the dynamic range of the output signal; and

means to apply to the modified components the inverse of the said wavelet transform, to produce an output signal.

Other, preferred, aspects of the invention are defined in the claims.

Some embodiments of the invention will now be described, by way of example, with reference to the accompanying drawings, in which:

Figure 1 is a block diagram of one form of speech processing apparatus according to the invention;

Figures 2 and 3 are a block diagram of two possible implementations of the wavelet transform unit of Figure 1;

Figures 4 and 5 are block diagrams of two possible implementations of the inverse transform;

Figures 6a and 6b show graphically two versions of the Daubechies wavelet;

Figure 7 is a graph of a test speech waveform;

Figures 8 and 9 are graphs showing respectively the transformed version of the test waveform and the clipped test waveform;

Figure 10 shows one implementation of the processing unit in Figure 1;

Figure 11 is a graphical representation of a test waveform and a clipped test waveform after processing by the apparatus of Figure 1; and

Figures 12 to 14 show some alternative wavelets.

The apparatus of Figure 1 is designed to receive, at an input 1, speech signals which have been distorted by clipping. The input signals are assumed to be in the form of digital samples at some sampling rate f_s, e.g. 8 kHz. On the assumption that, because of the clipping, the signal employs the whole of the available dynamic range of the digital representation, it is firstly multiplied, in a multiplier 2, by a scaling factor S_1 ($S_1 < 1$) to allow "headroom" for subsequent processing. Of course, an analogue-to-digital converter may be added if an analogue input is required. The signals are then supplied to a filter arrangement 3 which applies to the signals a Wavelet Transform, to produce N (e.g. five) outputs corresponding to respective transform levels. The series of signals n_i ($i = 1, \ldots N$) appearing at these outputs are fed to a processing unit 4 which scales or otherwise processes them to produce N processed outputs n_i' which are then subject to the inverse wavelet transform in an inverse transform unit 5, to provide, after further scaling by a multiplier 6, a reconstructed speech signal at an output 7.

The general form of the wavelet transform W_g of a function f(t) is

$$W_g = \int f(\tau)\, g\left(\frac{\tau - b}{a}\right) d\tau$$

where g is the transform kernel.

If b is regarded as the independent variable of W_g, and expressed as a time series for discrete values a_i of a, then (writing also a summation for the integral as we are dealing with a discrete system) we have a set of series for the transformed signal:

$$W_{a_i} = \sum_{\tau = t-(n-1)}^{t} f(\tau)\, g\left(\frac{\tau - t}{a_i}\right)$$

where i (i = 1,...N) is the level of the series and n is the number of filter coefficients.

If we write $g(x/a_i) = g_i(x)$ then

$$\nu_i = W_{a_i} = \sum_{\tau=t-(n-1)}^{t} f(\tau)\ g_i(\tau - t)$$

which can be implemented by a bank of N filters having coefficients given by g_1 to g_N respectively. Such a filter bank is shown in figure 2 with N filters 31/1 to 31/N.

The transform kernel g can in principle be any wavelet, i.e. a temporally finite waveform having a mean value of zero; however, particularly preferred is the use of a Daubechies wavelet, a formal definition of which may be found in I Daubechies "Orthonormal Bases of Compactly Supported Wavelets", Comm. Pure & Applied Maths, Vol. XLI, No. 7, pp 909-996 (1988), incorporated herein by reference.

In this embodiment, $a_i = 2^i$.

Because of the limited bandwidth of the filters their outputs contain a lot of redundant information and can be downsampled to a lower sampling rate by decimators 32/1 to 32/N, in each case by a factor $k_i = 2^i$.

Alternatively, the filter bank may be constructed from cascaded quadrature mirror filter pairs, as shown in Figure 3, where a first pair 33/1, 34/1 with coefficients g and h feed decimators 35/1, 36/1 (of factor 2) and so on. Comparison with Figure 1 shows that $h = g_1$. Note that, unlike the Figure 2 construction, this structure has a further output, referenced 37 in Figure 3, carrying a residual signal - i.e. that part of the input information not represented by the N transformed outputs. This may be connected directly to the corresponding input of the synthesis filter.

Figure 4 shows one implementation of the inverse transform unit 5, with upsampling devices 51/1, 51/2 ... 51/N having the same factors $k_1...k_N$ as the decimators in Figure 2, followed by filters 52/1, 52/2, ... 52/N having coefficient sets g1', g2', ... gN' whose outputs are combined in an adder 53. Each coefficient set g1′ etc.. is a time-reversed version of the coefficient set g1 etc.. used for the corresponding filter in Figure 2.

Figure 5 shows a cascaded quadrature mirror filter form of the inverse transform unit 5, with filters 54/1, 54/2, ... 54/N having coefficients h′ and filters 55/1, 55/2, ...55/N with coefficients g′. h′ and g′ are time-reversed versions of the coefficient sets h and g respectively, used in Figure 3. Upsamplers 56/1, 56/2, ... 56/N and 57/1, 57/2, ... 57/N are shown, as are adders 58/1, 58/2, ... 58/N. Each section is similar; for example the second section receives the second order input, upsamples it by a factor of two in the upsampler 56/2 and passes it to the filter 54/2. The filter output is added in the adder 58/2 with the sum of higher-order contributions fed to the second input of the adder via the x2 upsampler 57/2 and filter 55/2. The highest order section receives the residual signal at its upsampler 57/N. The output of the unit 5 is produces by the adder 58/1.

As an analysis method wavelet transforms are, ideally, characterised by the qualities of completeness of representation, which implies invertability, and orthogonality, which implies minimal representational redundancy. Furthermore, in principle, one could adopt the notion that the mother wavelet (or wavelets) should be designed to closely match the characteristics of speech such that the representation is compact, in the sense that as few coefficients as possible in the transform domain have significance.

The Daubechies wavelet transform has neatly rounded triangle of orthogonality, scale and translation factors and invertability. The cost is that the waves are completely specified and are therefore generic and cannot be adapted for speech or any other signal in particular.

Now it may be that for power of two decimations figure 3 is actually a general form and that the Daubechies theory actually amounts to the imposition of orthogonality and invertibility with this.

We can see the shape of the Daubechies wavelets by direct analysis of the structure in Figure 3 to obtain the equivalent filters of Figure 1. For the fourth order transform the first dilated wavelet is 10 samples long, the second 22, the third 46; for the sixth order these numbers are 16, 36 and 76. A direct numerical method to get these is to inverse transform impulses at and scale level. This was done to obtain figures 6a and 6b showing, respectively, a 6th -rder, 4th level Daubechies wavelet and a 20th order 4th level Daubechies wavelet; where the discrete Fourier transform of the wavelets is also shown beneath. It is seen that they are band limited signals and that the lower order wavelets have significant ripple.

The effect of clipping in the wavelet transform domain is illustrated by Figures 7 to 9. Figure 7 shows a test waveform of 0.5 seconds of speech, plotted against sample number at 8 kHz. Figure 8a - 8e show the 12^{th} order Daubechies wavelet transform of the test waveform, to five levels, plotted against sample number after decimation, whilst Figure 9a-e shows the same transform of the test waveform clipped at ± 1000 (referred to the arbitrary vertical scales on Figure 7). Figures 8f and 9f show the residual signal in each case.

The task of the sequence processor 4 is to process the sequences of Figure 8a-e such that they more closely resemble those of Figure 9a-e. The simplest form of this processing is a linear scaling of the sequences, and the version shown in Figure 10 shows multipliers 41/1 etc. applying the following factors:

first level	0.2
second level	0.2
third level	0.68
fourth level1	
fifth level	1

This arrangement acts to rebuild the dynamic range of the signal by enhancing the longer scale components of the Wavelet transform, since it was observed that these are apparently only scaled by clipping. The final scale factor s2 should be chosen by some AGC method.

Figure 11 shows a sample a of speech and b of the same speech after clipping and processing by the apparatus of Figure 1, with the weights given above, S1 = 1 and S2 = 2.5. Clip levels are marked CL.

The determination of the best weights more formally can be done if a cost function can be defined. Some experiments involving manual search were performed using dynamic range matrixes of peaks characterised by the median of the top five absolute values in a sample, and troughs characterised by the number of samples of value less than 5.

For practical implementation the best weights should be determined from direct numerical optimisation.

Other forms of processing are possible, for example a nonlinear scaling, of possibly forming linear or nonlinear combinations of sequences. Nonlinear operations may include thresholding, windowing, limiting and rank order filtering.

If desired, this weighting may be adaptively controlled. Two aspects are addressed here.

Firstly the clipping levels may change. We might assume that s1 in some way tracks this, that is to say s1 = 1 when no clipping is present and decreases otherwise. Then if we were using fixed weights, W^{o}, determined by some one-time optimisation we might use

$$W = s_1 + (1 - s_1)W^{o}$$

in the filter or something more complicated.

Secondly the off-line weight determination may not be adequate for the range of speech signal actually occurring on the line. In that case it could be advantageous to adaptively alter the weights in real time. At present there is no analytic cost of the weight available. A numerical function could be the product of the dynamic range measures discussed above. Since there are only a few weights in the wavelet domain filter it is feasible to do a direct gradient search. Exploring all possibilities of adding or subtracting a given step to each weight involves the evaluation of the cost function $2^n + 1$ times for n weights (the number of vertices of an n-dimensional hypercube plus one for the centre point). This can be implemented by providing this number of filters with the appropriate shifted weight vectors and replacing the centre value with best performing one at set time steps.

The Wavelet Domain Filter based on the Daubechies sequence works very well. The Daubechies wavelets is generic and one might expect that better results could be obtained with wavelets that are closely matched to the speech signals themselves. In doing this it would be expected that use can be made of the fact that voiced speech is more likely to suffer from clipping. That is to say the wavelet series can, in principle, be tailored to represent in a compact and thus easily processed form, the parts of speech sensitive to clipping.

The main problem here is the design of the wavelet transform, the mother wavelet and the set of scaling and translation to be employed and how they are implemented.

In designing matched wavelets it will be very difficult to retain the orthogonality and perfect reconstruction properties that the Daubechies transform has. We will need to understand the trade-off between these properties and the improved representational powers of the sophisticated wavelets. It may be that appropriate orthogonality and slightly imperfect reconstruction would be sufficient if there were clear gains in the representational power.

There has been some work on fitting mother wavelet shapes in a least squares sense in order to achieve improved data compression. One seeks to parameterise the shape of the wavelets in some way and perform direct optimisation. Here the zero crossing patterns are used to find wavelets for the filter bank structure; only the first level is considered. Examining the zero-crossing statistics of the test waveform shows that there are repeated patterns of two or more components. The general form of most of these is a "down-chirp"; large followed by smaller intervals. As a simple ad-hoc way of building wavelets with given zero crossing intervals, parabolas were joined together. Some wavelets designed this way are shown in Figures 12 to 14.

CLAIMS

1. An apparatus for processing speech comprising:
means to apply to a speech signal a wavelet transform to generate a plurality of transformed components;
means to modify the component such as to increase the dynamic range of the output signal; and
means to apply to the modified components the inverse of the said wavelet transform, to produce an output signal.

2. An apparatus according to Claim 1 in which the transform is a Daubechies wavelet transform.

3. An apparatus according to Claim 2 including decimators for reducing the sampling rate of the components prior to modification.

4. An apparatus according to Claim 3 in which the transform means is formed by cascaded quadrature mirror filter pairs.

5. An apparatus according to one of the preceding claims in which the modifying means is operable to apply weighting factors to at least some of the components.

6. An apparatus according to Claim 5 in which the weighting factors are relatively lower for relatively lower order components.

7. An apparatus according to Claim 5 or 6 including means for measuring the degree of clipping of the speech signal and to vary the weighting factors as a function thereof.

8. Apparatus for processing speech substantially as herein described with reference to the accompanying drawings.

Fig.1.

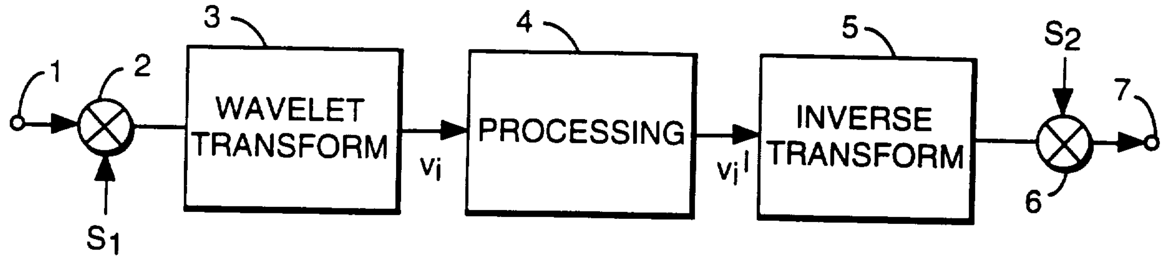


Fig.4.

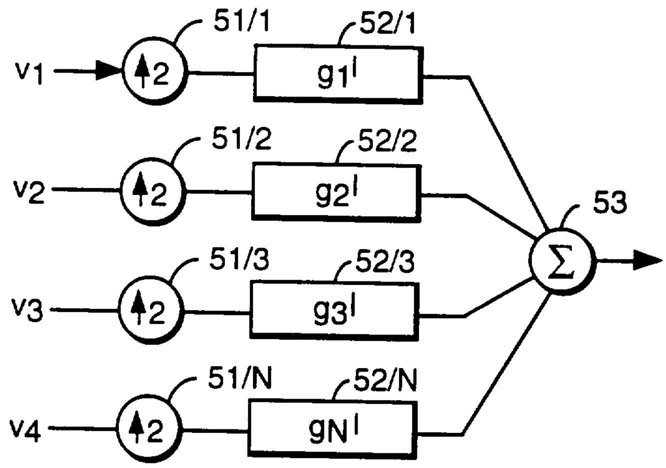


Fig.10.

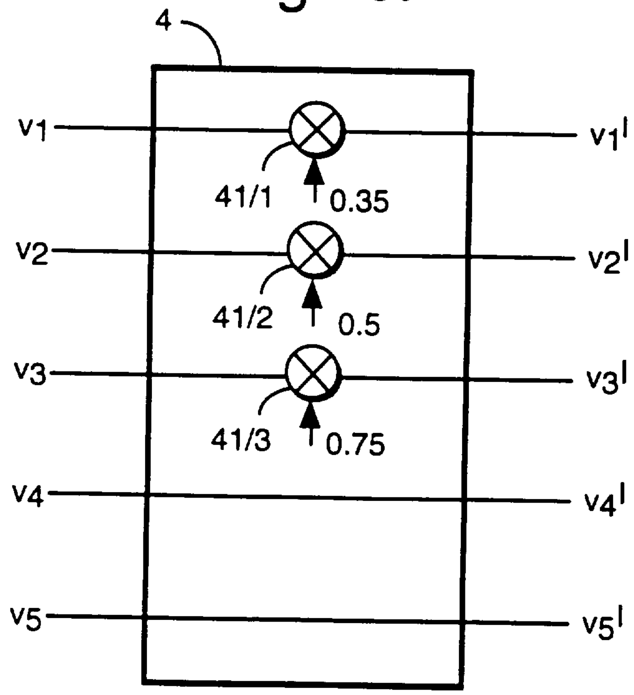


Fig.5.

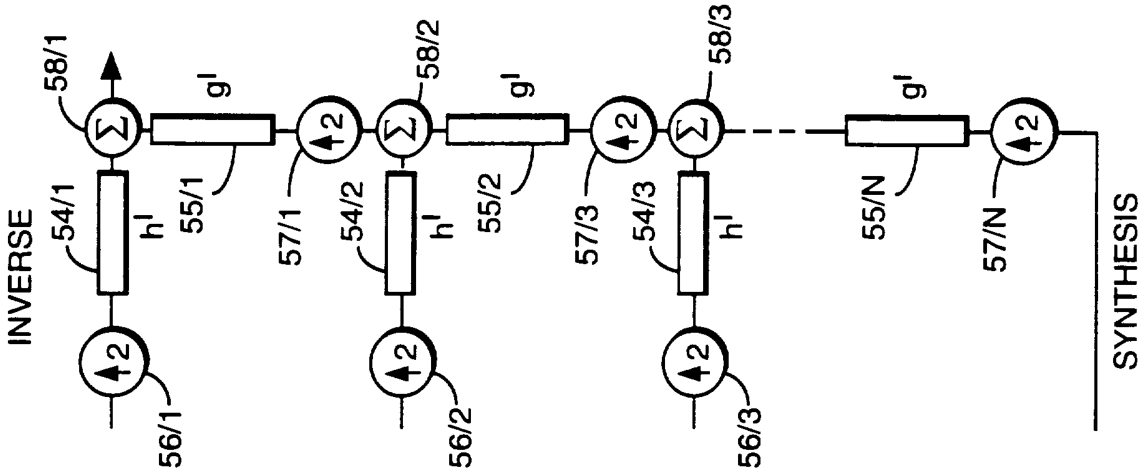


Fig.3.

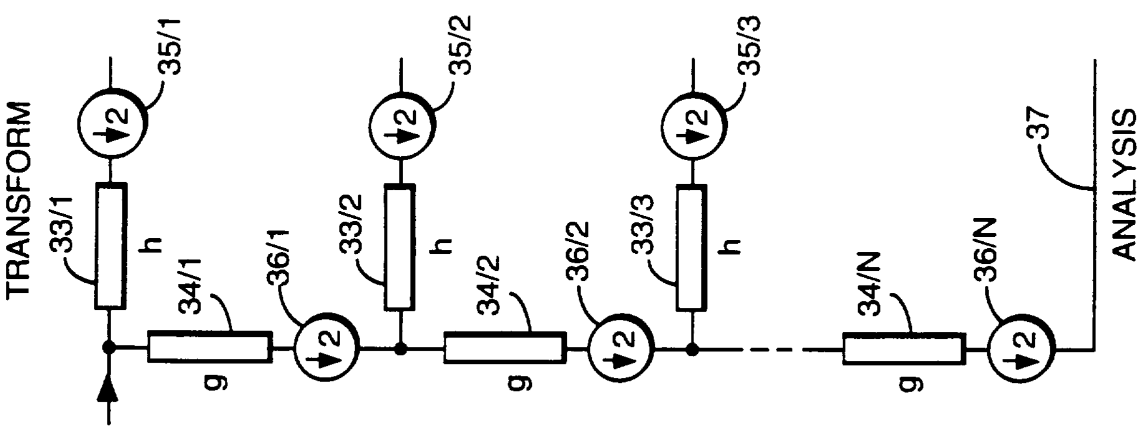


Fig.2.

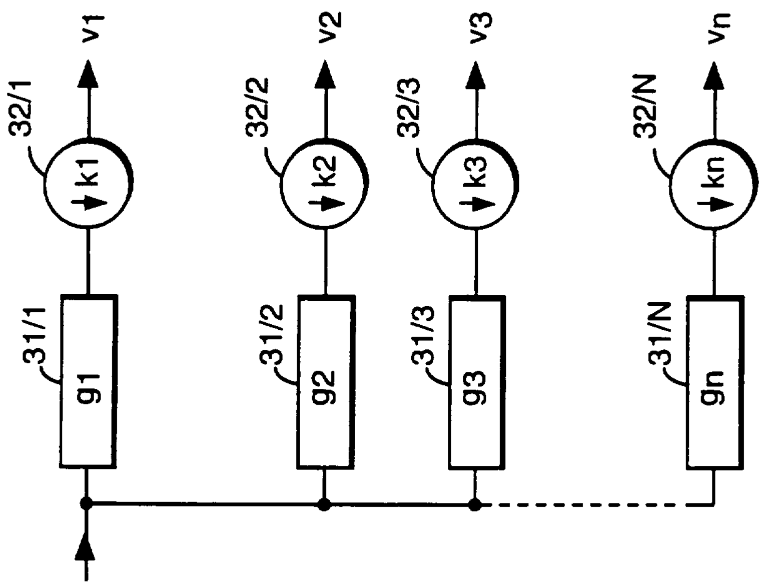


Fig.6b.

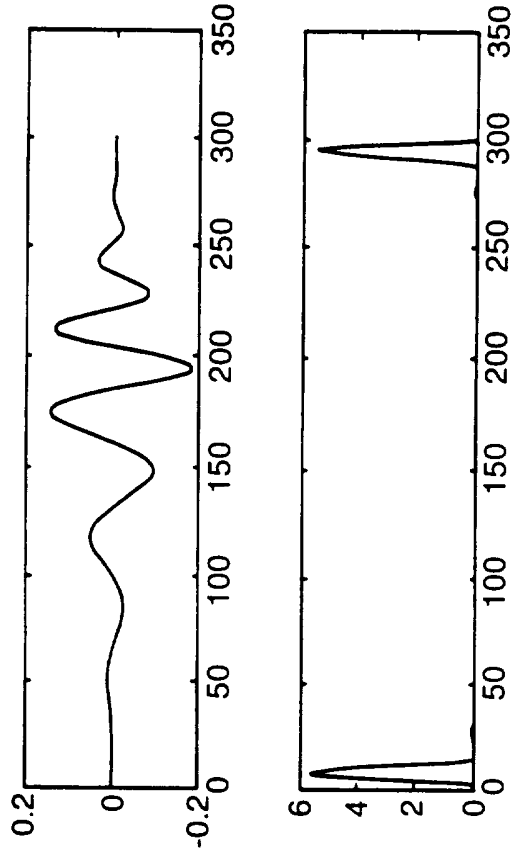

0.2
0
-0.2
0 50 100 150 200 250 300 350
6
4
2
0
0 50 100 150 200 250 300 350

Fig.6a.

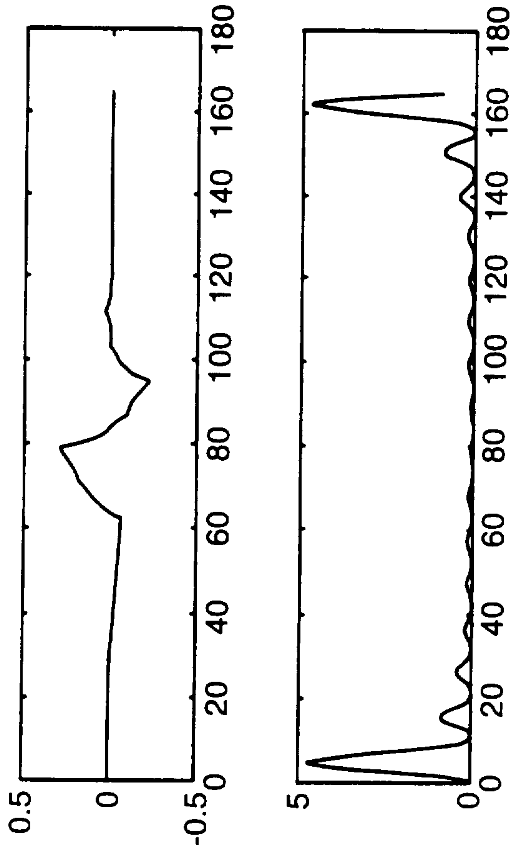

0.5
0
-0.5
0 20 40 60 80 100 120 140 160 180
5
0
0 20 40 60 80 100 120 140 160 180

Fig.7.

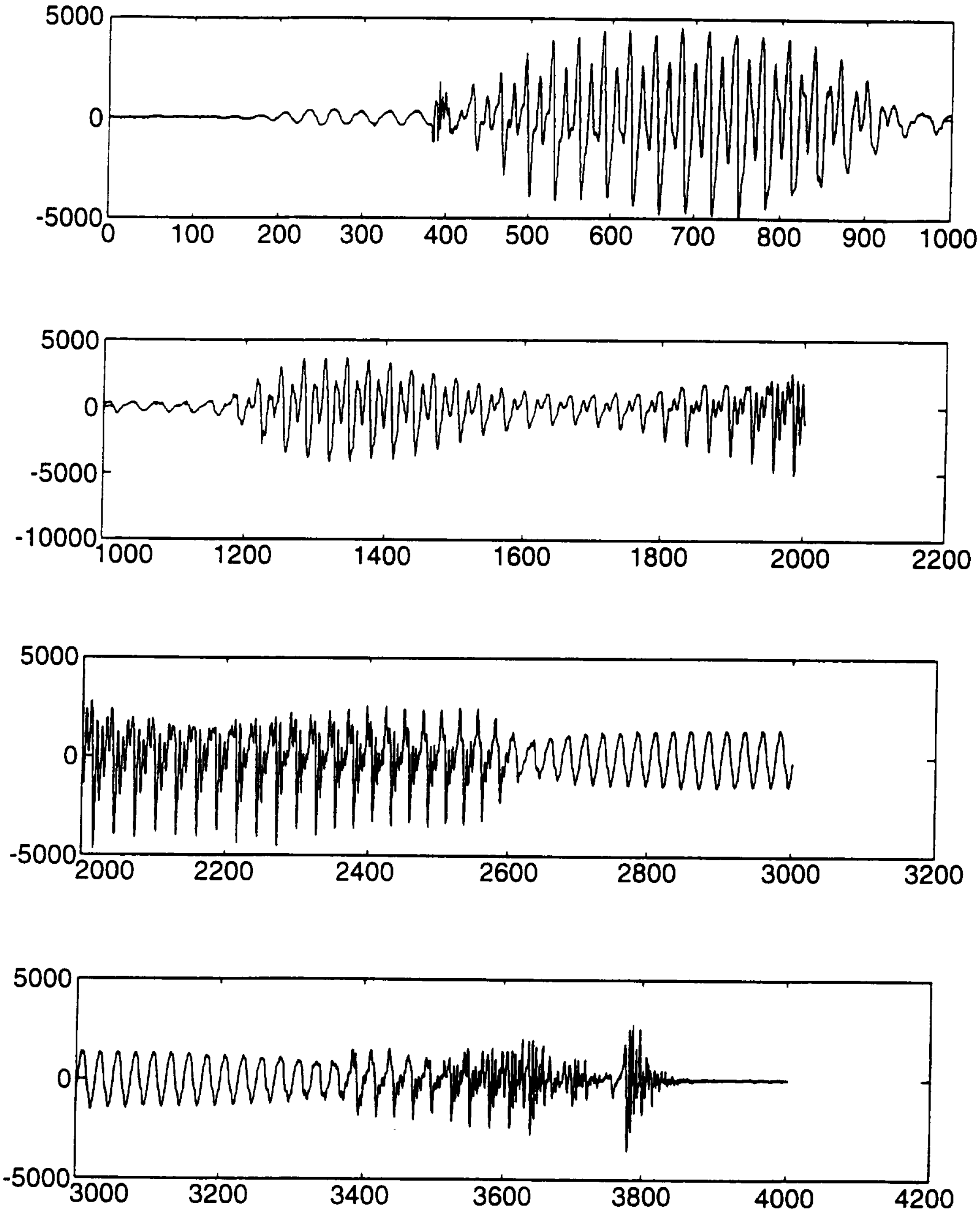

5000
0
-5000
0
100
200
300
400
500
600
700
800
900
1000
5000
0
-5000
-10000
1000
1200
1400
1600
1800
2000
2200
5000
0
-5000
2000
2200
2400
2600
2800
3000
3200
5000
0
-5000
3000
3200
3400
3600
3800
4000
4200

Fig.8.

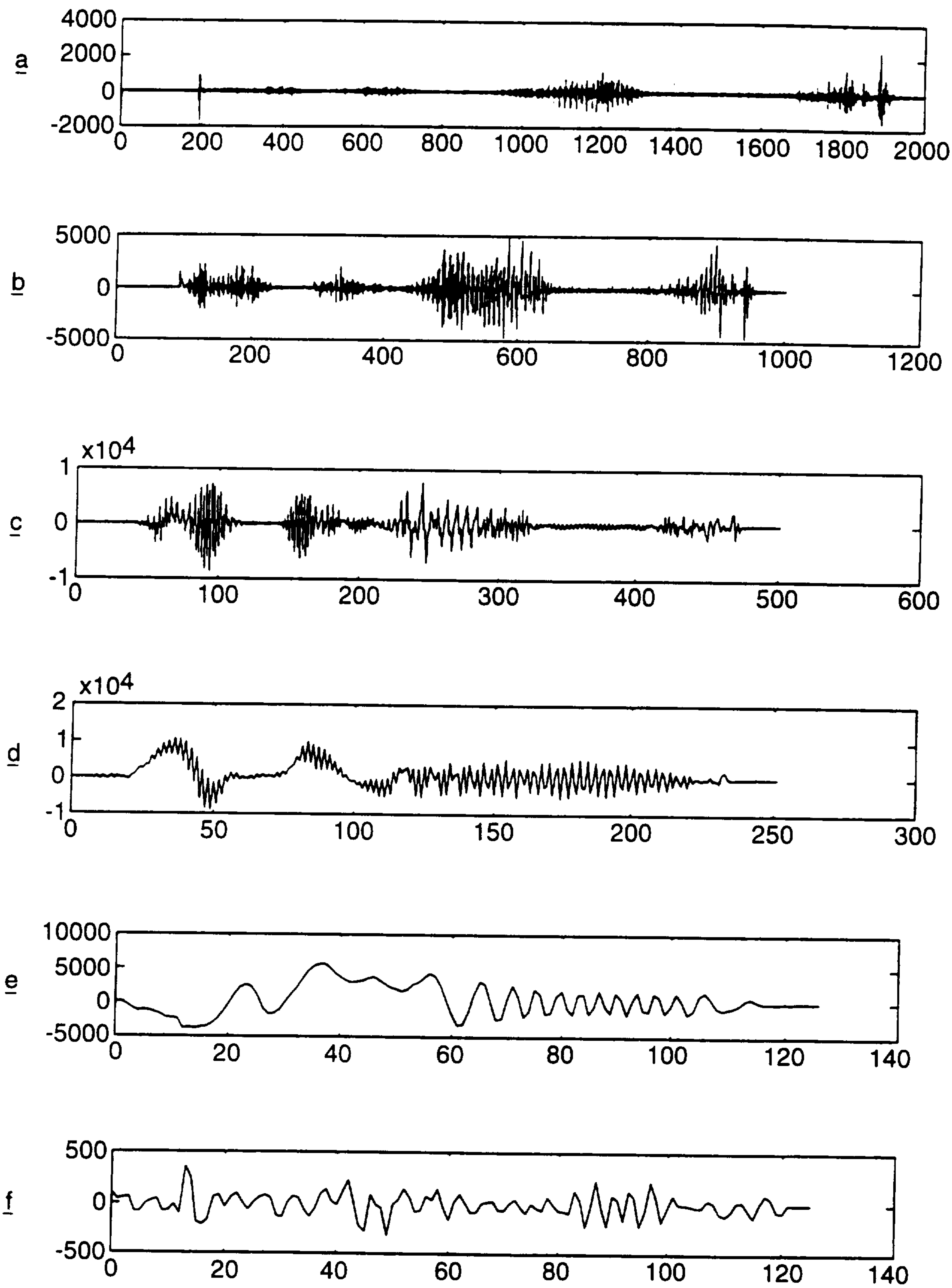

a
4000
2000
0
-2000
0 200 400 600 800 1000 1200 1400 1600 1800 2000
b
5000
0
-5000
0 200 400 600 800 1000 1200
c
x10⁴
1
0
-1
0 100 200 300 400 500 600
d
x10⁴
2
1
0
-1
0 50 100 150 200 250 300
e
10000
5000
0
-5000
0 20 40 60 80 100 120 140
f
500
0
-500
0 20 40 60 80 100 120 140

Fig.9.

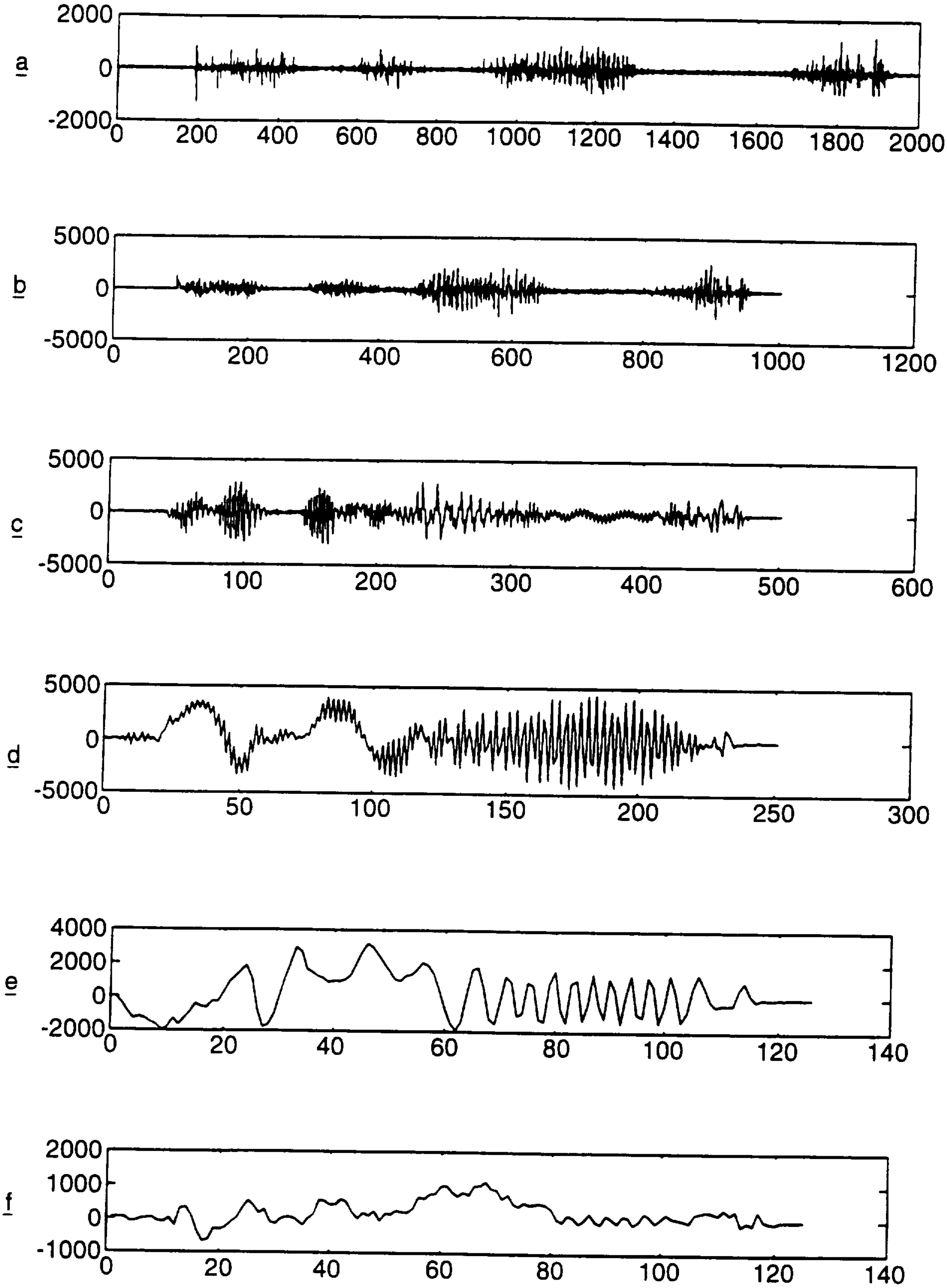

a
2000
0
-2000
0 200 400 600 800 1000 1200 1400 1600 1800 2000
b
5000
0
-5000
0 200 400 600 800 1000 1200
c
5000
0
-5000
0 100 200 300 400 500 600
d
5000
0
-5000
0 50 100 150 200 250 300
e
4000
2000
0
-2000
0 20 40 60 80 100 120 140
f
2000
1000
0
-1000
0 20 40 60 80 100 120 140

Fig.11.

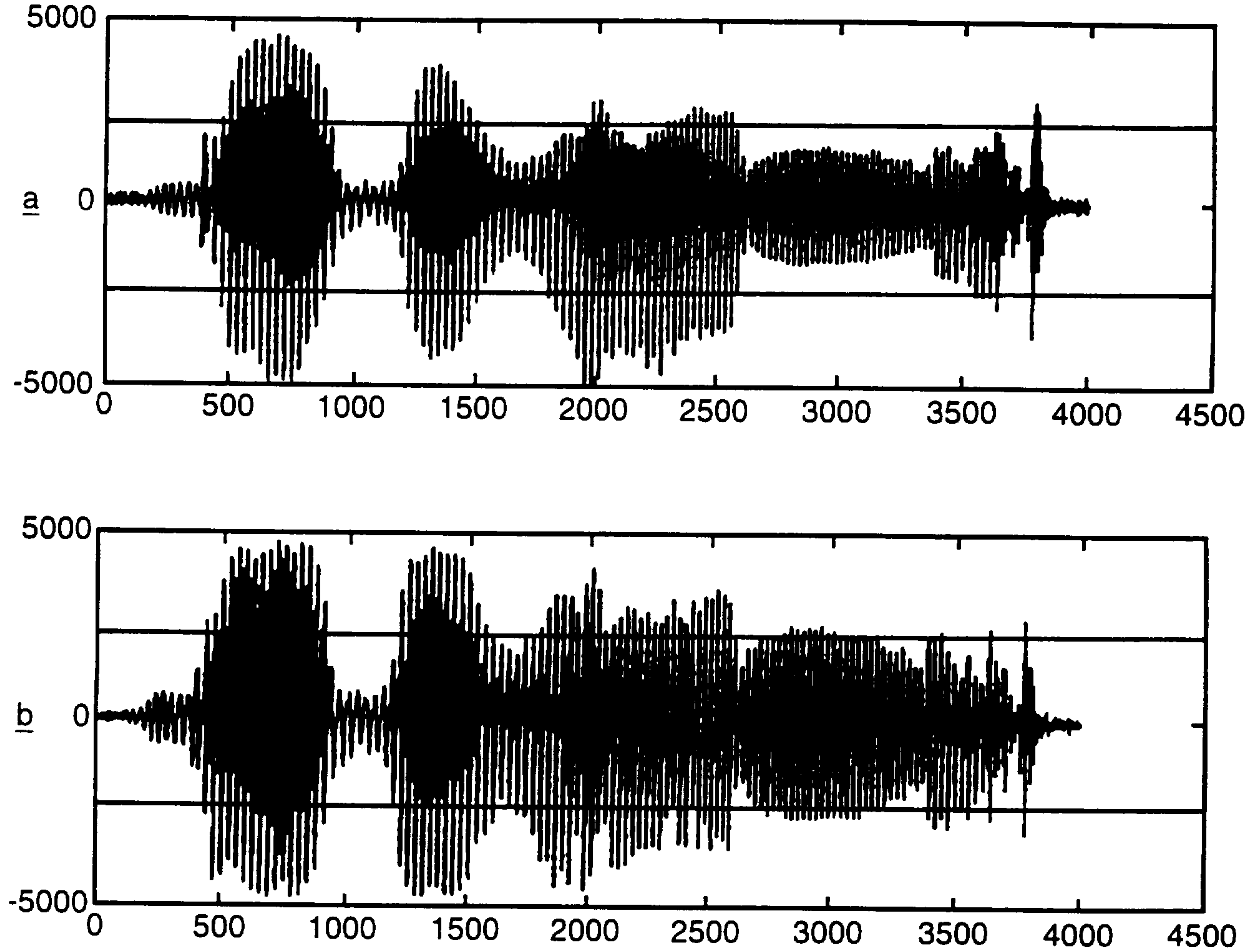

5000
a 0
-5000
0 500 1000 1500 2000 2500 3000 3500 4000 4500
5000
b 0
-5000
0 500 1000 1500 2000 2500 3000 3500 4000 4500

Fig.12

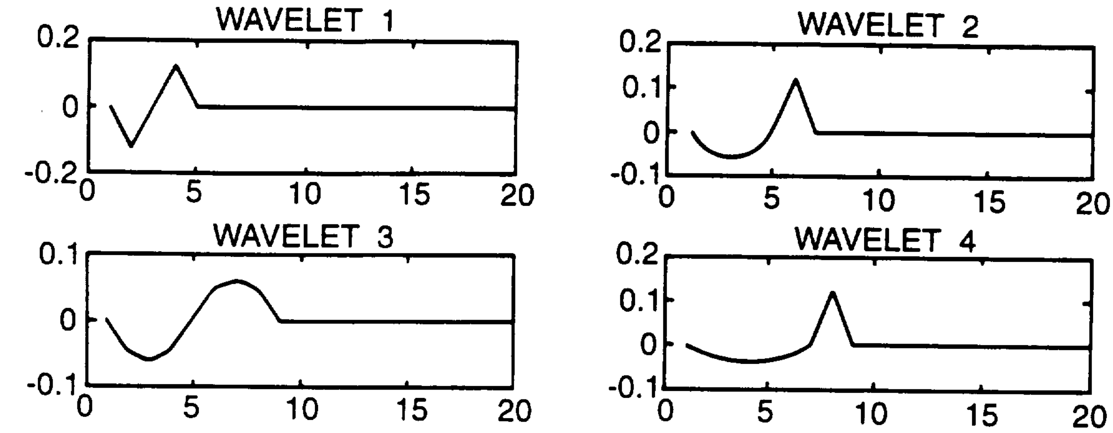


Fig.13

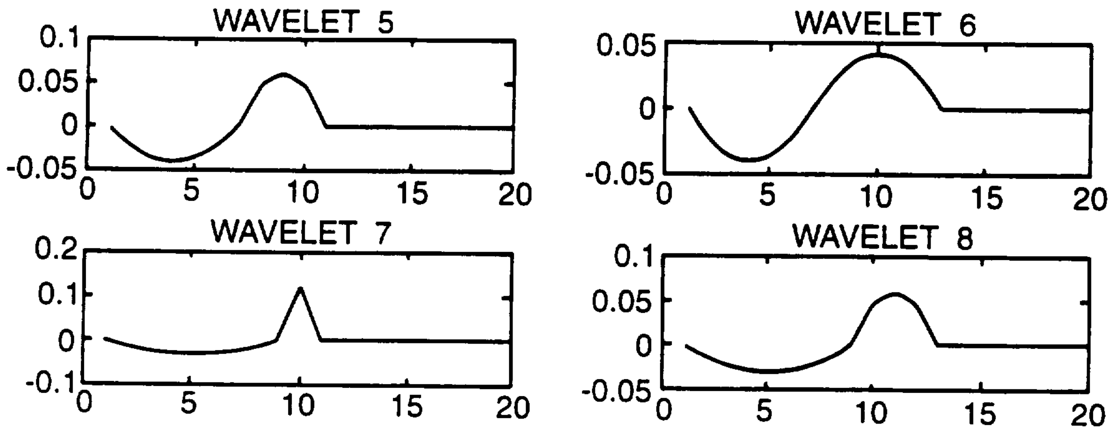
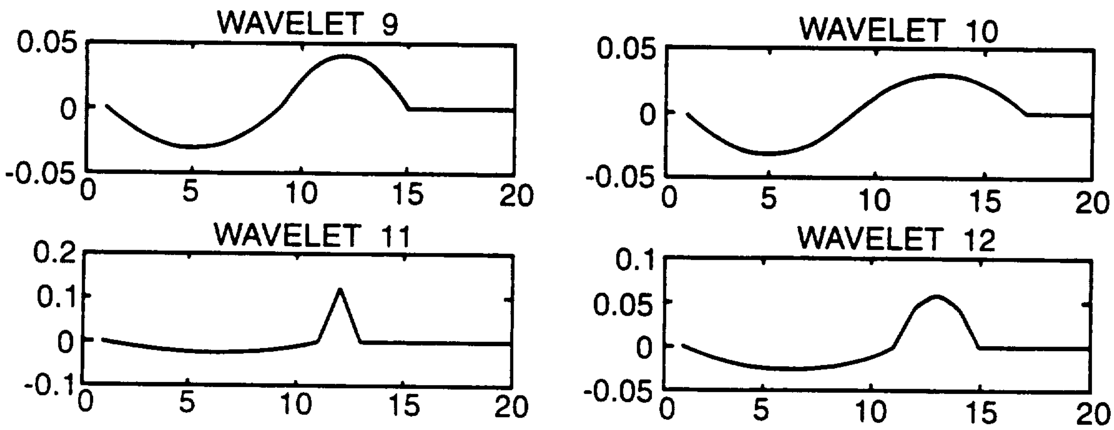


Fig.14



INTERNATIONAL SEARCH REPORT

Inter   onal Application No
PCT/GB 95/02943

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 G10L9/16

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 G10L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	ICASSP 92: 1992 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING (CAT. NO.92CH3103-9), SAN FRANCISCO, CA, US, 23-26 MARCH 1992, ISBN 0-7803-0532-9, 1992, NEW YORK, NY, US, XP002000224 IRINO T ET AL: "Signal reconstruction from modified wavelet transform-An application to auditory signal processing" see page 85 - page 88 see page 85, paragraph 2. - page 85, paragraph 3. ---	1
A	WO,A,84 02992 (AURETINA PATENT MANAGEMENT) 2 August 1984 see page 4, line 23 - page 6, line 15	1
A	see page 68, line 6 - line 20 ---	7

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
29 March 1996	10.04.96
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax (+31-70) 340-3016	Authorized officer Lange, J

INTERNATIONAL SEARCH REPORT

International Application No
PCT/GB 95/02943

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US,A,4 974 187 (LAWTON) 27 November 1990 see column 2, line 29 - line 44 see column 18, line 34 - column 19, line 2; figure 12	1
A	see column 1, line 40 - line 64 ---	4
A	WO,A,89 06877 (BRITISH TELECOM) 27 July 1989 see page 14, line 18 - page 18, line 10; figures 7,8 ---	1
A	EP,A,0 621 582 (MATRA COMMUNICATION) 26 October 1994 see page 1, line 27 - line 41 -----	1

INTERNATIONAL SEARCH REPORT

Inter   nal Application No
PCT/GB 95/02943

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO-A-8402992	02-08-84	US-A-	4559602	17-12-85
		AU-B-	2497284	15-08-84
		CA-A-	1217279	27-01-87
		CA-C-	1231454	12-01-88
		EP-A-	0134238	20-03-85
		JP-T-	60500275	28-02-85
US-A-4974187	27-11-90	AU-B-	6069790	11-03-91
		EP-A-	0485444	20-05-92
		JP-T-	4507315	17-12-92
		WO-A-	9102317	21-02-91
WO-A-8906877	27-07-89	CA-A-	1332626	18-10-94
		DE-D-	68913139	24-03-94
		DE-T-	68913139	23-06-94
		EP-A-	0367803	16-05-90
		GB-A-	2220330	04-01-90
		JP-T-	2503256	04-10-90
		US-A-	5133013	21-07-92
EP-A-0621582	26-10-94	FR-A-	2704348	28-10-94
		FI-A-	941874	24-10-94
		JP-A-	7121197	12-05-95