



(12)

EUROPEAN PATENT APPLICATION

(21) Application number : **91301034.4**

(51) Int. Cl.⁵ : **G10L 9/02**

(22) Date of filing : **08.02.91**

(30) Priority : **08.02.90 GB 9002852**

(43) Date of publication of application :
14.08.91 Bulletin 91/33

(84) Designated Contracting States :
DE FR GB IT SE

(71) Applicant : **NATIONAL RESEARCH
DEVELOPMENT CORPORATION
101 Newington Causeway
London SE1 6BU (GB)**

(72) Inventor : **Holmes, John Nicholas
19 Maylands Drive
Uxbridge, Middlesex UB8 1BH (GB)**

(74) Representative : **Hasler, Christopher
Patent Department National Research
Development Corporation 101 Newington
Causeway
London SE1 6BU (GB)**

(54) **Methods and apparatus for spectral analysis.**

(57) In automatic speech recognition it is usual to make a spectral analysis of the incoming speech signal and it can be useful to detect the frequencies and intensities of the formants. However although the formants are mostly quite well defined during vowel sounds there are frequent occasions when this is not so and it is not so during a high proportion of consonant sounds. The present invention determines the frequencies at which the centroids of respective frequency versus power distributions occur in a plurality of frequency bands of a signal representing speech (approximately corresponding to the ranges of individual formants). The centroids have most of the desirable properties of formants but also carry significant information for those sounds for which the conventional definition of formants does not seem appropriate. Preferably the powers in the bands in which the centroids are measured are also determined. The incoming signal is filtered (2) into separate frequency bands and the power in each band is measured (4). The output signal in each band is weighted by 3 dB per octave (5) and then the power in that band is measured (6). The power ratio obtained (7) for a band from the power after 3 dB weighting divided by the power before weighting gives an indication of the position of the centroid of that band in the frequency spectrum.

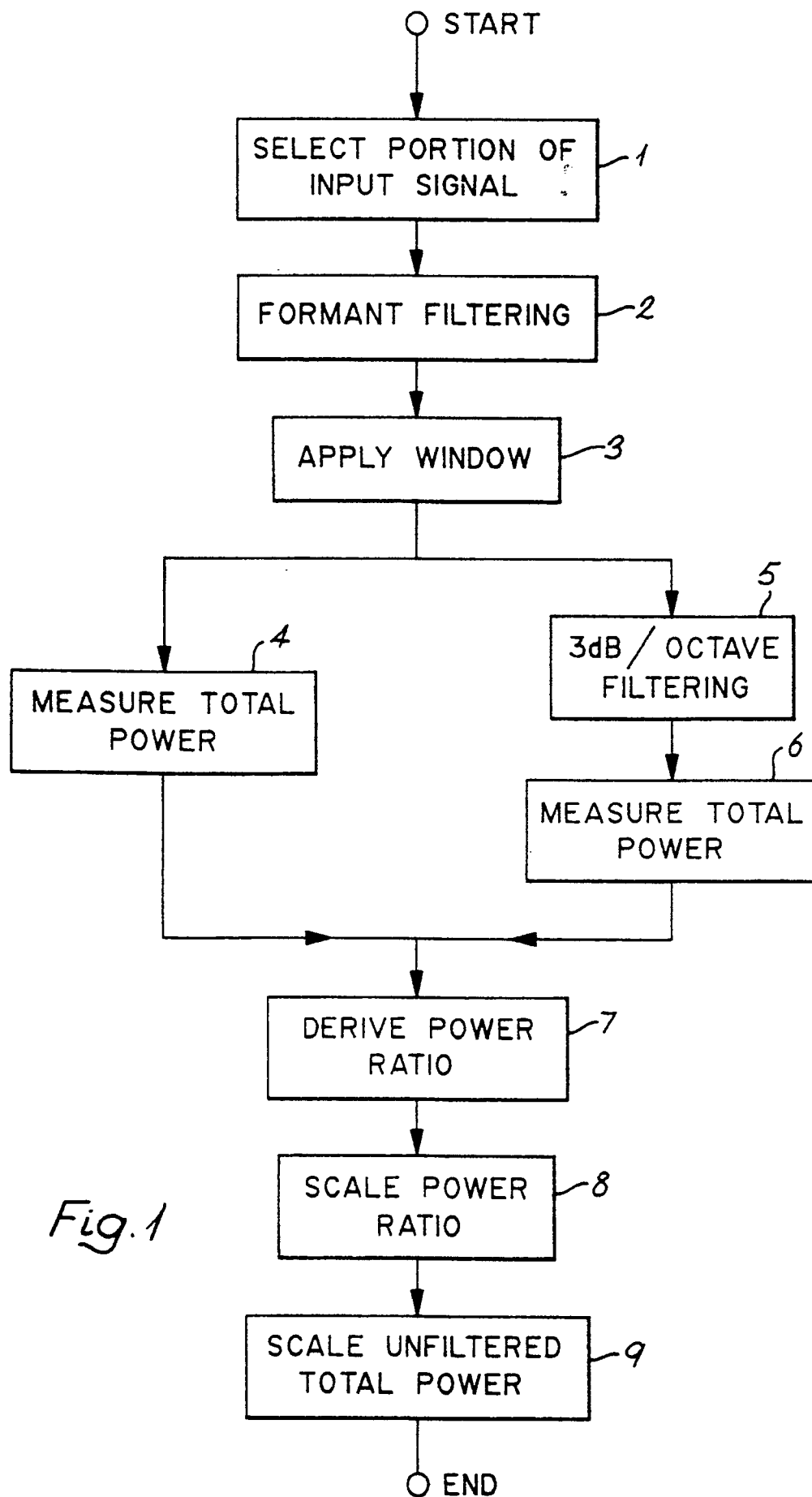


Fig. 1

METHODS AND APPARATUS FOR SPECTRAL ANALYSIS

The present invention relates to methods and apparatus for spectral analysis, particularly the spectral analysis of sounds produced in speech. Such analysis finds applications in, for example, automatic speech recognition and speech coding for bandwidth reduction and the storage of speech.

In automatic speech recognition it is usual to make a preliminary acoustic analysis of the speech signal to derive a description of the spectrum shape properties at regular intervals, typically in the range of 10 to 30 ms. The sets of measurements so derived are often referred to as "feature vectors", where each feature vector may typically contain between 5 and 20 features, depending on the method of analysis adopted. It is well known that the frequencies and intensities of the main short-term concentrations of power in a speech signal (formants) are highly correlated with the phonetic realization of the associated speech sound. The frequencies and intensities of the small number of formants that occur within the most significant part of the speech spectrum are useful as features for speech recognition. However, although during vowel sounds the formants are mostly quite well defined, there are also frequent occasions when study of a speech signal by spectrographic analysis fails to reveal clear formants, particularly during nasalization, during very weak vowels and during a high proportion of consonant sounds. It is therefore notoriously difficult to devise algorithms that will give robust formant frequency measurements during these less tractable sounds. The most reliable algorithms that have been published have required an excessive amount of computation to implement.

An important object of the present invention is to provide a set of features that have most of the desirable properties of formants, but also carry phonetically significant information even for those sounds for which the conventional definition of formants does not seem appropriate. A further object is to provide methods and apparatus for calculating these features easily.

According to a first aspect of the present invention there is provided a method for use in speech recognition of determining short-term characteristic features of a first signal representative of a speech signal, comprising the steps of

filtering the first signal to obtain time-varying second signals each in one of a plurality of frequency bands, and

determining at least approximate indications of the frequencies at which the centroids of respective frequency versus power distributions in the said bands occur as the characteristic features.

Preferably the powers in the bands in which the said centroids are measured are also determined as

further characteristic features.

As applied to speech sounds the first signal can be regarded as an electrical signal representative of a speech sound. The filtering to obtain further time varying signals is electrical filtering carried out for example by filters constructed from discrete components or by digital filters implemented by a computer such as a microprocessor. Although embodiments of the invention are described below in terms of digital computation applied to digitized sampled-data signals the invention may alternatively be implemented by analogue techniques.

The standard method of calculating the centroid of a distribution is to take the ratio of two integrals. If the distribution is represented graphically, the numerator of this ratio is the integral of the product of the ordinate and the abscissa, whereas the denominator is the integral of the ordinate. For spectral analysis these quantities refer to measurements in the frequency domain; the denominator integral is the total power in the duration of signal that is being analysed, which is the same in time domain as in the frequency domain and so can be computed in the time domain by summing the squares of the signal waveform samples. The numerator represents the sum of the powers of all spectral components after each component is multiplied by a quantity proportional to frequency. Weighting the amplitude of each spectral component by the square root of its frequency and then summing the squares of all the weighted spectral components would give the required numerator. Thus the numerator can also be computed in the time domain by passing the waveform samples through a filter whose gain is proportional to the square root of frequency over the relevant band, and squaring and summing the filtered waveform samples. The filter gain characteristic required has a positive slope of three dB per octave and can be approximated very closely by a sampled data filter of moderate order, using standard filter design methods. The power of each frequency band is given by the denominator integral.

Thus in the first aspect of the invention, for each frequency band, the step of determining at least approximations to the frequencies at which the centroids occur may comprise, for each filter output,

summing the squares of samples of the time varying signals at the filter output to provide a denominator which indicates the power of that filter output,

applying at least an approximation to a three dB frequency weight per octave to the samples,

summing the squares of the resultant samples to provide a numerator, and

dividing the numerator by the denominator to indicate the frequency of the centroid.

The invention also includes apparatus for carrying out the first aspect thereof.

The method of finding the frequency at which a centroid occurs from signals in the time domain can be generally applied. Therefore according to a second aspect of the invention there is provided

a method of determining short-term characteristic features of a first signal having a time-varying value comprising the steps of

filtering the first signal to obtain second time-varying signals each in one of a plurality of frequency bands, and

determining at least approximate indications of the frequencies at which the centroids of respective frequency versus power distributions in the said bands occur as the characteristic features by, for each frequency band,

determining the total power of the second signal for that band in the time domain to provide a first power value,

applying spectral weighting to frequency components of the second signal for that band,

determining the total power of the spectrally weighted signal in the time domain to provide a second power value, and

dividing the second power value by the first to provide an indication of the frequency of the centroid of that band.

According to a third aspect of the present invention there is provided

apparatus for determining short-term characteristic features of a signal having a time-varying value comprising

means for filtering a first signal having a time-varying value to obtain second time-varying signals each in one of a plurality of frequency bands, and

means for determining at least approximate indications of the frequencies at which the centroids of the frequency versus power distributions in the said bands occur as the characteristic features by,

determining the total power of the second signal for that band in the time domain to provide a first power value,

applying spectral weighting to frequency components of the second signal for that band,

determining the total power of the spectrally weighted signal in the time domain to provide a second power value, and

dividing the second power value by the first to provide an indication of the frequency of the centroid of that band.

In the second and third aspects of the invention, the spectral weighting may be at least an approximation to three dB per octave.

Instead of using a filter to apply the three dB frequency weighting per octave, the signal from the filter may be differentiated which, as is well known, is equivalent to applying a 6 dB per octave increase. If

an approximation to differentiation is carried out on a waveform represented by samples, by subtracting each sample from the previous sample, the increase is about 6 dB per octave at low frequencies, gradually reducing to zero slope as the half-sampling rate frequency is reached. The effects of the variation from the ideal 3 dB per octave slope are two-fold. First, the signals at higher and lower frequencies than the spectral peak are not given the correct relative weight in the centroid calculation. However, as these components are normally much weaker than the components near the spectral peak, this error makes very little difference to the measured peak frequency. If the signal being measured were a pure sinusoid there would be no error from this cause. Secondly, the calculated ratio of the powers will no longer be linearly related to the frequency of the spectral peak. This non-linearity need cause no problem, because the measured value can be converted to linear frequency by a pre-computed look-up table. For frequencies near to half the sampling rate, where the frequency-domain slope of the differencing operation tends to zero, there would be almost no sensitivity to frequency change, so this method is not suitable for use in this range. This problem can be avoided by spectrally inverting the signal before taking measurements in the upper half of the frequency range, so bringing high frequencies down near to zero, where differencing is effective.

For speech recognition and speech coding applications, the filtering of the time varying signals can be carried out by any bandpass filters which correspond approximately to the ranges of the three lower formants; typically 250 to 900 Hz, 700 to 3,000 Hz and 1,800 to 3,500 Hz respectively. However as will be described later some shaping of the filter characteristics is preferable in order to separate the two formants when they occur in overlapping parts of the filter characteristics.

It is well known in signal processing theory that when a spectrum measurement of a short duration of signal is made there are errors in the resultant spectrum caused by the relation between the signal waveform shape and the precise points in time over which the signal is analysed. The usual method to reduce these end effects is to multiply the signal by a smooth function of time known as a "window", which attenuates those parts of the signal near the ends of the analysis period. The same considerations apply to spectral analysis by centroid measurement, so that more consistent results are obtained if the signal is windowed before the centroid measurement. Thus a time-window may be applied to the band limited signal at the output of each filter and then followed by "three dB per octave" filtering and summation operations applied to the entire duration of the windowed output. The accuracy of this process can be ensured by using a finite-impulse-response "three dB per octave" filter,

so that it is known that the output is exactly zero once the impulse response duration has passed.

An informal explanation of why the centroid of power works fairly well as a feature in speech recognition is now given. The intensity of any formant peak is usually several dB above the spectral intensity elsewhere in the band allocated to that formant. If the power spectral density remote from the formant is, say, 14 dB below that round the formant peak (which would be quite normal), this figure would represent a power ratio of 25 ; even if the remainder of the band covered by the filter were five times wider than that of the high intensity part around the formant peak, it would still only contribute a fifth of the total power, and so would not disturb the centroid much. To a large extent such disturbance would, in any case, be systematic, always biasing the formant estimate slightly towards the centre of the band-pass filter. Any such systematic variation could be corrected by subsequently applying a non-linear function to the formant measurement, but in practice it would not matter for any type of pattern-matching speech recognizer because the same systematic effect would apply similarly during the recognizer's training process. An advantage of using the centroid instead of direct measurement of the peak frequency is that it will always give an unambiguous result even when the spectral peak or peaks are not clearly defined. Provided the same method of analysis is used for setting up patterns for the pattern-matching process in the recognition algorithm, the fact that the measurements do not always correspond to formants is not important.

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

Figure 1 is a flow diagram of an algorithm according to the invention for finding the frequency of the centroid of a formant and the power in the band in which the centroid is measured,

Figure 2 is a typical windowing shape used in the algorithm of Figure 1,

Figure 3 shows a typical first formant waveform in the time domain,

Figure 4 shows a typical spectrum of the waveform of Figure 3,

Figure 5 is a flow diagram of a first part of a microprocessor algorithm for an 8-bit microprocessor, Figure 6 is a flow diagram of a second part of the microprocessor algorithm approximating to the second and third operations of Figure 1 as applied to the first formant,

Figures 7a and 7b form a flow diagram of a third part of the microprocessor algorithm for measuring total power in a formant band and for applying an approximate 3 dB per octave filter to the band before deriving the total power of the resultant output,

Figure 8 is a flow diagram of a fourth part of the

microprocessor algorithm equivalent to the last three operations of Figure 1, and

Figure 9 is a block diagram including apparatus according to the invention.

Speech signals from a microphone are converted to a linear digital representation by a suitable A-D conversion system sampling at 8 kHz. Preliminary audio spectral shaping and gain control is provided such that the full range of the A-D conversion system is used and there is a good average balance between high and low frequency components of the signal. The shaping and gain control is also arranged to attenuate the low frequency prominence normally occurring below the frequency of the first formant during voiced sounds.

The output of the A-D conversion system is connected to a computer such as a digital signal processor (DSP) integrated circuit or a microprocessor which in speech recognition carries out the recognition algorithms in addition to feature extraction based on centroids. A general algorithm (see Figure 1) is first described and then followed by a description of more specific algorithms for use with an 8-bit microprocessor.

The sampled signal at the output of the A-D conversion system may be divided in frames each containing a predetermined number of samples. In an operation 1 of Figure 1 a portion of each frame of duration of, for example, 2 to 30 ms is selected, longer durations (even up to a complete frame) being preferred if sufficient computation power for the centroid measuring process is available.

Samples from operation 1 are digitally filtered in an operation 2 in three pass bands to obtain three groups of samples relating to the three lowest formants. Figure 1 shows the processing of only one of these formants and therefore in the complete analysis algorithm, all the operations of Figure 1 following operation 2 are repeated for the other two formants.

The formant filtering of operation 2 may be carried out by any suitable method. The pass bands for the filters have already been mentioned but for some speech sounds two formants can sometimes move very close together in frequency, and fall within the pass bands of two filters. It is a consequence of the acoustic theory of speech production that when two formants are close they are usually of approximately equal intensity, and so any centroid measurement in these cases is likely to give the mean of the two frequencies of the formants within the band instead of the one formant that was desired. This error can be appreciably reduced if the formant band filters are designed to have a sloping characteristic of, say, about 3 dB per 100 Hz in the overlap regions. As it is rare for two formants to be closer than about 300 Hz this means that their relative intensities will be very different, thus ensuring that the weaker one in each band would contribute very little to the centroid calcu-

lation. The resultant error in the formant intensity measurement is corrected by applying the inverse of the filter characteristic to the intensity result, as a function of the measured frequency.

As has been mentioned, when spectrum measurement of short duration signals is carried out it is desirable to apply a window which attenuates those parts of the signal near the ends of the analysis period. This process is carried out in operation 3 where the samples at the output of the formant filter selection are multiplied by the characteristic of Figure 2 over the interval in which they occur.

The next step in finding the frequency of the centroid of the selected formant band is to measure the total power in the signal at the output of the window. In the time domain this signal is typically as shown in Figure 3 and has the distribution of Figure 4 in the frequency domain. Operations 4 to 7 of Figure 1 have the object of finding a denominator and a numerator from time domain signals as already outlined. The denominator is the power in the waveform of Figure 3 while the numerator is found by applying a spectral weighting in the form of a gain characteristic with a positive slope of 3 dB per octave to samples of the waveform of Figure 3 (the operation 5), measuring the total power in the resultant waveform (the operation 6) and dividing the output of the operation 6 by the output of the operation 4 to derive a power ratio (the operation 7) representative of the frequency of the centroid.

In an operation 8 the power ratio from the operation 7 is multiplied by a scale factor to convert to formant frequency, and in the operation 9 a scaled logarithm of the unfiltered power from the operation 4 is calculated to represent formant intensity in dB.

Describing now the use of an 8-bit microprocessor in the invention, a type 6502 for example, running at 4 MHz may be used for the recognition of continuous speech with a fairly small recognition vocabulary if the techniques described below are used to simplify multiplication and division during feature extraction, and very efficient computational techniques (not relevant to the present invention) are used for recognition.

For an 8-bit microprocessor and 8 kHz sampling, the input signal may be divided into frames containing 256 samples, that is 31.25 frames per second. However, with such a microprocessor limitation of computational power means that a detailed analysis to determine the formant centroids can only be carried out on a selected part of each frame.

It is the object of the operation 1 of Figure 1 as applied to 8-bit microprocessors to ensure that this part of the waveform includes the higher intensity parts of the signal for analysis (for example just after glottal closure for voiced sounds, and at stop bursts and the more intense parts of fricatives). The microprocessor program firstly compares every sample value in a frame with the largest sample previously

found in that frame in order to determine the largest sample in the frame (operation 11, Figure 5). Having found the largest sample, 20 sample portions which include the largest sample are tested in order to determine which such portion contains the maximum power (operation 12). The start point of a 20 sample window is moved over the 20 samples before the point of maximum amplitude. Having set the start point the powers in the next twenty samples are added and the power of the earliest sample is repeatedly discarded while the power of one new sample is added until the window has been moved by 20 samples. While the movement is in progress the maximum of the power in the window is stored together with the sample index at which it occurs. There is a potential problem with dynamic range in this integration of powers, because the 8-bit samples would need a 16-bit range if squared. Also, the time to do the squaring operation directly is not available. These problems are overcome by using a look-up table to give a suitably scaled version of the squares, and to select one of five differently-scaled tables according to the previously determined maximum level in the window. As the actual value of the sum of squares is not required, but merely the sample index at which it occurs, compensation for this scaling is not required later.

Since an interval longer than 20 samples is required for analysis, 50 samples starting with the 20 sample portion are selected for analysis in an operation 13.

In order to make best use of the 8-bit microprocessor an operation 14 is now carried out in which each sample is multiplied by the largest power of 2 that does not cause any samples of the fifty to exceed the range -128 to +127.

For 8-bit microprocessor operation filters for separation for the formant bands can be made using a cascade connection of simple finite impulse response (FIR) sections each with one or at the most two multiplication and addition operations. Within each section signal delays can be one or two sample periods or integer multiples of these numbers. Delays of two or more sample periods in one filter section imply multiple sets of zeroes in the transfer function, thus enabling higher order filters to be achieved without significantly increasing the computational load. To avoid the need for conventional multiplications the filter coefficients can be chosen to have values such as +1, -1, 0.5, 1.75, which can be implemented by at most a very small number of shift and add/subtract operations. For example the filter transfer function for the first formant filter may be

$$(1-Z^{-6})(1+Z^{-2})(1+Z^{-1})(1+Z^{-1}),$$

while that for the second formant filter may be $(1-Z^{-3})(1-Z^{-3})(1+Z^{-2}+Z^{-4})(1+1.75Z^{-2}+Z^{-4})(0.5+Z^{-2})$, where Z^{-1} denotes a delay of one sample interval.

Similar principles are used in designing the filter for the third formant. In general for economy of com-

putation the multiplying coefficients of the transfer functions of the sections are chosen to have the form $\pm 2^n(1 \pm 2^{-m})$ where n and m are integers, $0 \leq m \leq 3$ and $-1 \leq n \leq 1$.

Use of such simple filters does not provide such good formant separation as could be achieved by more conventional FIR designs, but the difficulty of designing flat filter characteristics within the computational constraints ensures that such filters intrinsically provide slopes in band overlap regions as suggested above. The errors in the intensity features caused by the lack of flat response in the main part of each pass band may, if required, be corrected by applying a look-up table depending on the result of the frequency measurement.

In Figure 6, representing the operation 2 of Figure 1, each sample is processed by the algorithm shown in order to separate the first formant. Similar algorithms are required to select the second and third formants but different transfer functions are employed. First the sample number is initialized to zero in an operation 16 and then the appropriate overall transfer function is achieved by applying the transfer functions of operations 17 to 20 in turn.

For the second and third formants an approximate version of the windowing of Figure 2 is carried out by simply halving the amplitude of the first and last samples in each measurement interval; this requires a single shift operation in each case. For the first formant, however, there may possibly be only one or two cycles of the formant frequency within the measurement interval and such a simple technique would therefore not be sufficient to prevent serious errors from end effects. A practical alternative which is implemented by operations and tests 22 to 28 in Figure 6 is to choose the two ends of the measurement interval to be at zero crossings of the signal and thus for the first formant operations 22 to 28 replace the windowing operation 3 of Figure 1. The first ten samples are disregarded with respect to zero crossing detection to exclude the initial transient of the first formant filter which is tenth order and uses a maximum delay of ten sample intervals. Thus the test 22 allows the operation 23 to increment the sample number and input a new sample, if the sample number is less than 10. The test 24 is carried out if the sample number is greater than 10 and determines whether the polarity of the current sample is opposite to that of the previous sample; if not the operation 23 is carried out and the next sample is taken but if so then a test 25 is carried out to determine whether this is the first zero crossing. If it is then the sample number of this zero crossing is stored in the operation 26 and then the next sample is taken but if not then the test 27 is carried out to determine whether more than 2 ms have elapsed since the first zero crossing. Thus if less than 2 ms have occurred then the next sample is taken but sampling ceases by the operation 28 if the output of

the test 27 is positive and the sample number of this final zero crossing is stored in an operation 28. Analysis of the output of the filtering operation 2 for the first formant therefore is applied only to the first sample following the first zero crossing after the first ten samples, the samples following in the next 2 ms and the following interval up to the next zero crossing.

The operations 4, 5 and 6 of Figure 1 as applied to an 8-bit microprocessor program are now described in more detail with reference to Figures 7a and 7b. For the reasons already given, this part of the algorithm uses logarithms for multiplication and division whereas for a DSP integrated circuit it is quite convenient to carry out these operations directly.

First an operation 30 of Figure 7a is carried out in which the sample number is initialized at the first sample number as stored in operation 26 of Figure 6. Two variables LSUMD and LSUMN are then initialised to zero in an operation 31 to represent eventually the logarithm of the total power in the interval and the logarithm of the total power after approximate 3 dB/octave filtering, respectively.

LSUMD is found by a process which includes finding an approximation to the logarithm of the sum of two numbers without using antilogarithms. This process is described by Kingsbury and Rayner (1971) "Digital Filtering Using Logarithmic Arithmetic", Electronic Letters, 7, pages 56 to 58 and in the inventor's book "Speech Synthesis and Recognition", published by Van Nostrand Reinhold (UK) Co. Ltd. in 1988, pages 149 and 150.

Kingsbury and Rayner pointed out that $\log(A+B) = \log(A(1+B/A)) = \log A + \log(1+B/A)$, and thus their process for finding the logarithm of the sum of two numbers A and B is as follows:

- (1) If $\log(B)$ is $> \log(A)$ then transpose $\log(A)$ and $\log(B)$,
- (2) Find $\log(B/A)$ by forming $\log(B) - \log(A)$,
- (3) Use the result of (2) to select a value from a look-up table, and
- (4) Add the result of (3) to $\log(A)$.

The look-up table is entered as $\log(B/A)$ and the table output is $\log(1+B/A)$, where A is the greater of two values: the power found up to the current sample; and the square of the current sample. B is the smaller of these two values.

In the present instance, as each sample is processed, the logarithm of the power up to that sample plus the square of the value of the present sample is found by the Kingsbury and Rayner process. Thus in an operation 32 of Figure 7a a look-up table is used to find the logarithm of the square of the current sample which is designated LSSAM. Then in a test 33 and an operation 34 the greater of LSSAM and the logarithm of the power found up to this sample (LSUMD) is found, the greater being designated LSUMD and the smaller LSSAM. In an operation 35, equivalent to finding $\log(B/A)$, the difference between

LSUMD and LSSAM is found and then, in order to find the term $\log(1+B/A)$ the look-up table mentioned above for the Kingsbury Rayner process is used in operation 36. Finally an operation 37 is carried out which gives the power up to and including the current sample by adding LSUMD to the result from the look-up table which is equivalent to finding $\log A + \log(1+B/A)$.

In Figure 7b, the approximate 3 dB/octave filtering of the operation 5 of Figure 1 is replaced by determining the difference between current and previous samples in an operation 38. It has already been explained that although at low frequencies this process gives a 6 dB/octave increase it can be used as an approximation.

In an operation 39 the logarithm of the square of each difference from the operation 38 is found by means of a look-up table and designated LSDIF. The logarithm of the sum of LSDIF and LSUMN is found using the Kingsbury and Rayner process in operations 40 to 44 in the same way as is described for LSSAM and LSUMD in the operations 33 to 37. Test 45 determines whether the last sample as stored in operation 28 of Figure 6 has been reached and if not a jump back to the operation 32 occurs and the next sample is taken. Otherwise operations 4, 5 and 6 have been carried out for a complete interval covered by the samples and an exit occurs from the algorithm of Figures 7a and 7b.

The equivalent algorithms to those of Figures 7a and 7b for the second and third formants initialise on the first sample of the interval and cease at the last sample without reference to zero crossings.

For the second formant the replacement for the approximate 3 dB/octave filtering is the differencing operation 38 but for the third formant where some frequency components may be near to half the sampling rate the frequency-domain slope of the differencing operation 39 tends to zero. Thus if this operation were used for the third formant there would be almost no sensitivity to frequency change and therefore the method is unsuitable for this formant. The problem is avoided by spectrally inverting the signal before taking measurements in the range of the third formant so bringing high frequencies down to near zero where the differencing is effective. The spectral inversion can be achieved by inverting every alternative waveform sample and the combined effects of spectral inversion and subsequent differencing are combined into the single operation of adding pairs of adjacent samples instead of differencing. However the frequency measurement so obtained then has to be subtracted from one half of the sampling frequency to compensate for the effects of spectral inversion. As far as Figure 7b is concerned all that is required is to change the operation 38 to one in which the sum of the current sample and the previous sample is determined, the subtraction from one half the sampling fre-

quency being carried out in an operation 48 of Figure 8.

The operation 7 of deriving the power ratio which gives the centroid frequency is now carried out by an operation 47 where the logarithm (LSUMD) of the denominator of the power ratio is subtracted from the logarithm of the numerator (LSUMN). The resulting value is then converted to a formant frequency by means of a further look-up table in the operation 48 which for the third formant includes subtraction of the frequency obtained from half the sampling frequency. The power in the band in which the centroid is measured is obtained in an operation 49 by a further look-up table which converts LSUMD to dB, taking account of the sloping characteristics of the formant filters in the overlap regions.

Thus as a result of carrying out the algorithm of Figure 1 or the approximations thereto of Figures 5 to 8 for each of the three formants, approximations to three formant frequencies and approximations to three formant powers are derived and can be used, for example, as features in speech recognition. Apparatus which includes the invention is shown in Figure 9 and comprises a signal capture portion 51 which includes the microphone, the audio spectral shaping and the A-D conversion system, a feature extraction portion 52 for carrying out the algorithm of Figure 1, or Figure 1 as approximated by Figures 5 to 8, and a pattern/modelling portion 53 for speech recognition from features obtained from the portion 52. The portions 52 and 53 are usually in the form of a single computer, DSP circuit or microprocessor as indicated above, which may also include some of the portion 51.

The invention can of course be put into operation in many other ways than those specifically described; for example a 16-bit or 32-bit microprocessor may be used and gives more accurate results since less approximations have to be made and larger signal portions are analysed. A DSP integrated circuit gives better results but may involve greater expense both in hardware and power consumption. Any other computers, apparatus or method for finding the centroids of spectral peaks can be used in spectral analysis according to the invention.

Claims

1. A method of determining short-term characteristic features of a first signal having a time-varying value comprising the steps of
 - filtering the first signal to obtain second time-varying signals each in one of a plurality of frequency bands, and
 - determining at least approximate indications of the frequencies at which the centroids of respective frequency versus power distribu-

- tions in the said bands occur as the characteristic features,
- characterized in that the said indications are determined by, for each frequency band,
- determining the total power of the second signal for that band in the time domain to provide a first power value,
- applying spectral weighting to frequency components of the second signal for that band,
- determining the total power of the spectrally weighted signal in the time domain to provide a second power value, and
- dividing the second power value by the first to provide an indication of the frequency of the centroid of that band.
2. A method according to Claim 1 characterized in that the spectral weighting is at least an approximation to three decibels per octave.
 3. A method according to Claim 1 or 2 characterized by including using the first power value for at least one of the frequency bands as a further characteristic feature of the first signal.
 4. A method according to Claim 1, 2 or 3 for spectral analysis of speech sounds characterized in that the frequency bands correspond to speech formants.
 5. A method according to any preceding claim characterized in that

the step of applying spectral weighting comprises differentiating, at least approximately, the second signal for that band.
 6. A method according to any preceding claim characterized by

deriving the first signal as a sequence of first samples by sampling an input signal at a predetermined rate, the step of filtering the first signal resulting in the second signals also being respective sequences of second samples, and

repeatedly deriving indications of the positions of the centroids of the distributions in the said frequency bands, each indication being derived from a group of successive second samples of one of the second signals.
 7. A method according to Claim 6, insofar as dependent on Claim 5, characterized in that

for each of the frequency bands in which the maximum frequency in the band is not near to the said rate, the step of applying the said spectral weighting is carried out by finding the difference in value between every sample and the previous sample in each group of second samples relating to that band to provide groups of successive difference samples.
 8. A method according to Claim 6 insofar as dependent on Claim 5, or Claim 7, characterized in that

for each of the frequency bands in which the maximum frequency in the band is near to the said rate, the step of applying the said spectral weighting is carried out by finding the sum of the values of each sample and the previous sample for every sample in each group of second signal samples relating to that band to provide groups of successive sum samples.
 9. A method according to Claim 6, 7 or 8 characterized in that the steps of determining the first and second power values comprise

taking each sample value of each group from which power values are to be derived in succession,

determining the logarithm of the square of the sample value so taken,

storing the logarithm of accumulated power in the sample powers up to the sample value taken,

determining which is the greater of the logarithm of the square of the sample value and the logarithm of the accumulated power,

subtracting the larger of these logarithms from the smaller to form the logarithm of the ratio of the greater of the sample power and the accumulated power, divided by the smaller,

determining the logarithm of one plus the said ratio by reference to stored values of the logarithm of the ratio versus the required logarithm,

adding the logarithm obtained from the stored values to the greater logarithm as previously determined to form the said logarithm of accumulated power,

whereby when each sample in a group has been taken the logarithm of accumulated power provides the power value required for that group of samples.
 10. A method according to Claim 9 characterized in that the step of dividing the second power value by the first comprises

subtracting the logarithm of the first power value from the logarithm of the second power value to provide the indication of the frequency at which the centroid occurs,

the logarithm of the first power value providing an indication of the power in the band having that centroid.
 11. A method according to any of Claims 6 to 10 characterized in that each group of successive samples of at least one of the second signals is

- derived by a windowing process comprising
 reducing the values of samples at the beginning and end of a succession of predetermined equal time intervals, each of which corresponds to one of the said groups, so that these samples are reduced in value by amounts which decrease from the beginning of each interval and increase with approach to the end of each interval.
12. A method according to any of Claims 6 to 11 characterized in that
 each group of successive samples of the second signals for the lowest frequency band are derived by a process comprising
 for each of a succession of equal predetermined time intervals,
 finding a first sample adjacent to a zero crossing in waveform samples from the lowest frequency band towards the beginning of each interval,
 finding a second sample adjacent to a zero crossing in waveform samples from the lowest frequency band towards the end of each interval,
 taking the samples of the second signals for each group as being delimited by the first and second samples of respective said time intervals.
13. A method according to Claim 6 characterized in that filtering the first signals comprises
 a plurality of finite impulse response filtering steps in cascade, the transfer functions of each step having multiplication coefficients of the form $\pm 2^n(1 \pm 2^{-m})$ where n and m are integers, $0 \leq m \leq 3$ and $-1 \leq n \leq 1$.
14. A method according to Claim 13 insofar as dependent on Claim 4 characterized in that
 at least two of the said frequency bands have overlapping regions and the frequency versus attenuation characteristics in the overlapping regions increase sufficiently towards the edges of the band to distinguish the formant corresponding to the band from a formant from an overlapping band.
15. Apparatus for determining short-term characteristic features of a signal having a time-varying value comprising
 means for filtering a first signal having a time-varying value to obtain second time-varying signals each in one of a plurality of frequency bands, and
 means for determining at least approximate indications of the frequencies at which the centroids of the frequency versus power distributions in the said bands occur as the characteristic features,
 characterized in that the said indications
- are determined by
 determining the total power of the second signal for that band in the time domain to provide a first power value,
 applying spectral weighting to frequency components of the second signal for that band,
 determining the total power of the spectrally weighted signal in the time domain to provide a second power value, and
 dividing the second power value by the first to provide an indication of the frequency of the centroid of that band.
16. Apparatus according to Claim 15 characterized in that the spectral weighting is at least an approximation to three decibels per octave.
17. Apparatus according to Claim 15 or 16 characterized by including using the first power value for at least one of the frequency bands as a further characteristic feature of the first signal.
18. Apparatus according to Claim 15, 16 or 17 for spectral analysis of speech sounds characterized in that the means for filtering is arranged to divide the first signal into frequency bands which correspond to speech formants.
19. Apparatus according to any of Claims 15 to 18 characterized in that the said means for determining is constructed to apply spectral weighting by differentiating, at least approximately, the second signal for that band.
20. Apparatus according to any of Claims 15 to 19 characterized by
 means for deriving the first signal as a sequence of first samples by sampling an input signal at a predetermined rate,
 the means for filtering the first signal having an output in the form of the second signals which are also respective sequences of second samples, and
 the means for determining being arranged to repeatedly derive indications of the positions of the centroids of the distributions in the said frequency bands, each indication being derived from a group of successive second samples of one of the second signals.
21. Apparatus according to Claim 20 characterized in that the means for determining is arranged to determine the first and second power values by
 taking each sample value of each group from which power values are to be derived in succession,
 determining the logarithm of the square of the sample value so taken,

storing the logarithm of accumulated power in the sample powers up to the sample value taken,

determining which is the greater of the logarithm of the square of the sample value and the logarithm of the accumulated power,

subtracting the larger of these logarithms from the smaller to form the logarithm of the ratio of the greater of the sample power and the accumulated power, divided by the smaller,

determining the logarithm of one plus the said ratio by reference to stored values of the logarithm of the ratio versus the required logarithm,

adding the logarithm obtained from the stored values to the greater logarithm as previously determined to form the said logarithm of accumulated power,

whereby when each sample in a group has been taken the logarithm of accumulated power provides the power value required for that group of samples.

- 22.** Apparatus according to Claim 21 characterized in that the means for determining is arranged to divide the second power value by the first by

subtracting the logarithm of the first power value from the logarithm of the second power value to provide the indication of the frequency at which the centroid occurs,

the logarithm of the first power value providing an indication of the power in the band having that centroid.

- 23.** Apparatus according to Claim 20 characterized in that the means for filtering the first signals comprises

a plurality of finite impulse response filtering means coupled in cascade, the transfer functions of each said filtering means having multiplication coefficients of the form $\pm 2^n(1 \pm 2^{-m})$ where n and m are integers, $0 \leq m \leq 3$ and $-1 \leq n \leq 1$.

- 24.** Apparatus according to Claim 23 characterized in that

the means for filtering is so arranged that at least two of the said frequency bands have overlapping regions and the frequency versus attenuation characteristics in the overlapping regions increase sufficiently towards the edges of the band to distinguish the formant corresponding to the band from a formant from an overlapping band.

- 25.** Apparatus constructed or arranged to carry out a method according to any of Claims 1 to 14.

- 26.** Apparatus according to Claim 15 comprising a computer or integrated circuit programmed to carry out at least one of the steps of any of Claims 1 to 14.

- 27.** A method for use in speech recognition of determining short-term characteristic features of a first signal representative of a speech signal, comprising the steps of

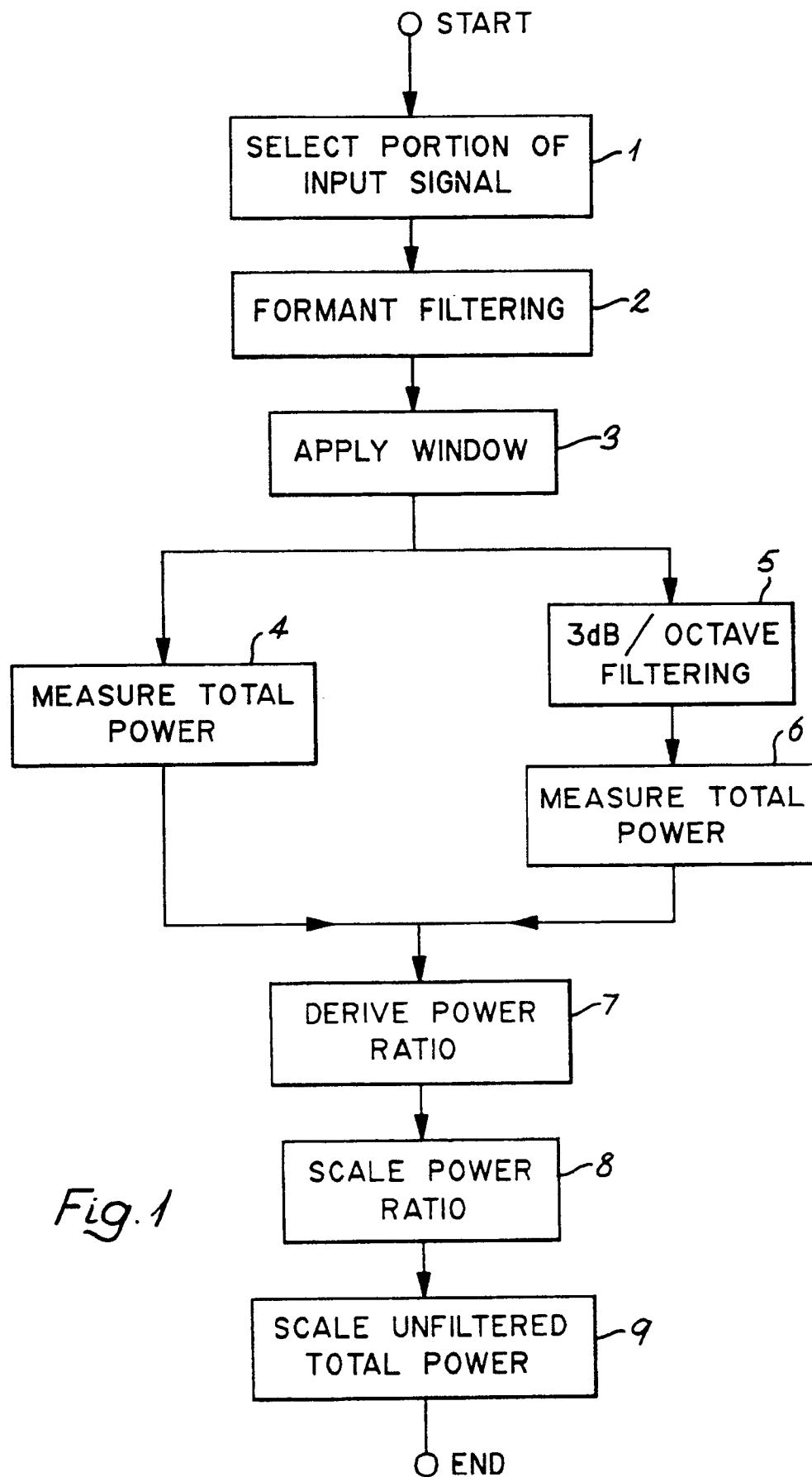
filtering the first signal to obtain time-varying second signals each in one of a plurality of frequency bands, and

determining at least approximate indications of the frequencies at which the centroids of respective frequency versus power distributions in the said bands occur as the characteristic features.

- 28.** Apparatus for use in speech recognition for determining short-term characteristic features of a first signal representative of a speech signal, comprising

means for filtering the first signal to obtain time-varying second signals each in one of a plurality of frequency bands, and

means for determining at least approximate indications of the frequencies at which the centroids of respective frequency versus power distributions in the said bands occur as the characteristic features.

*Fig. 1*

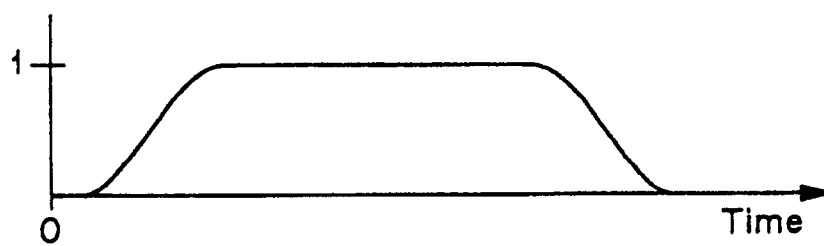


Fig. 2

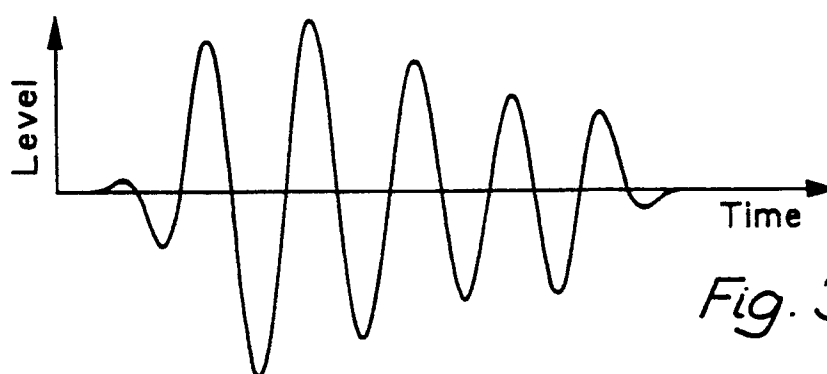


Fig. 3

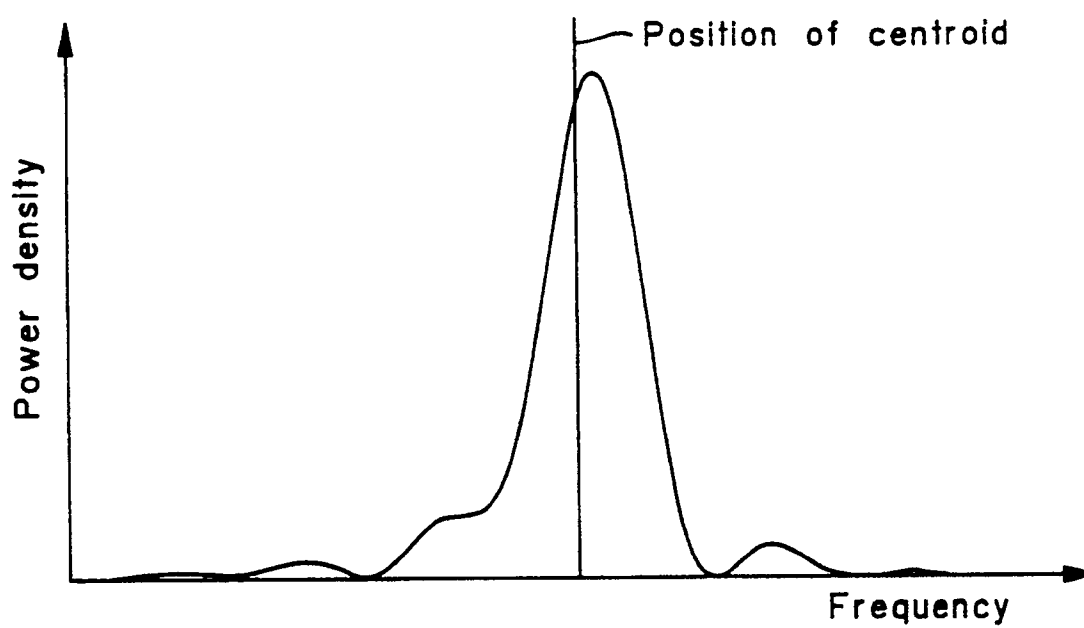


Fig. 4

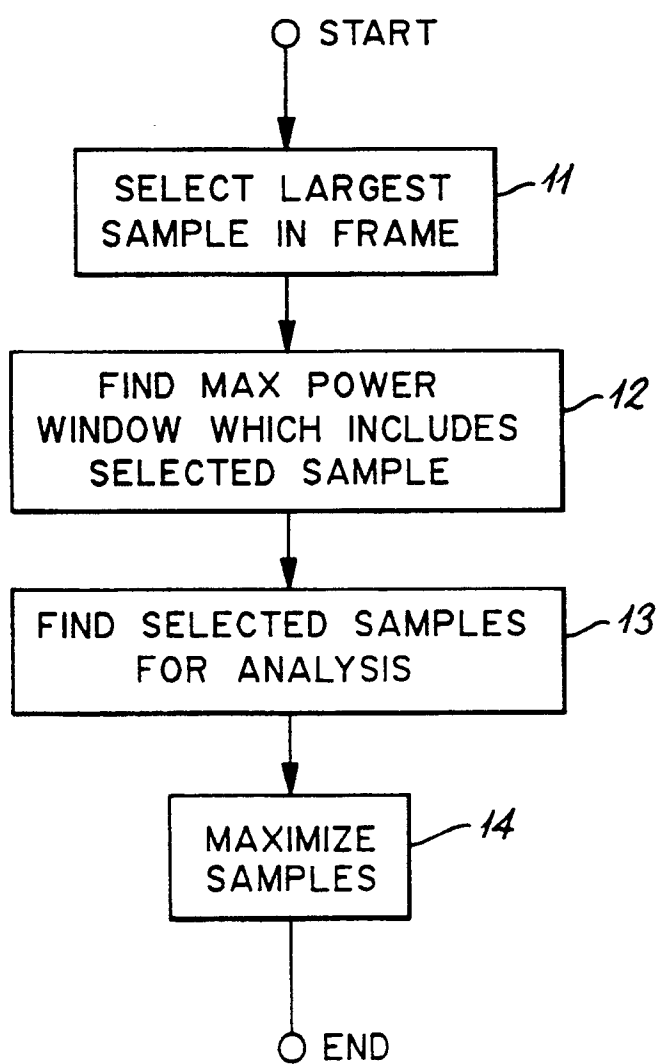


Fig. 5

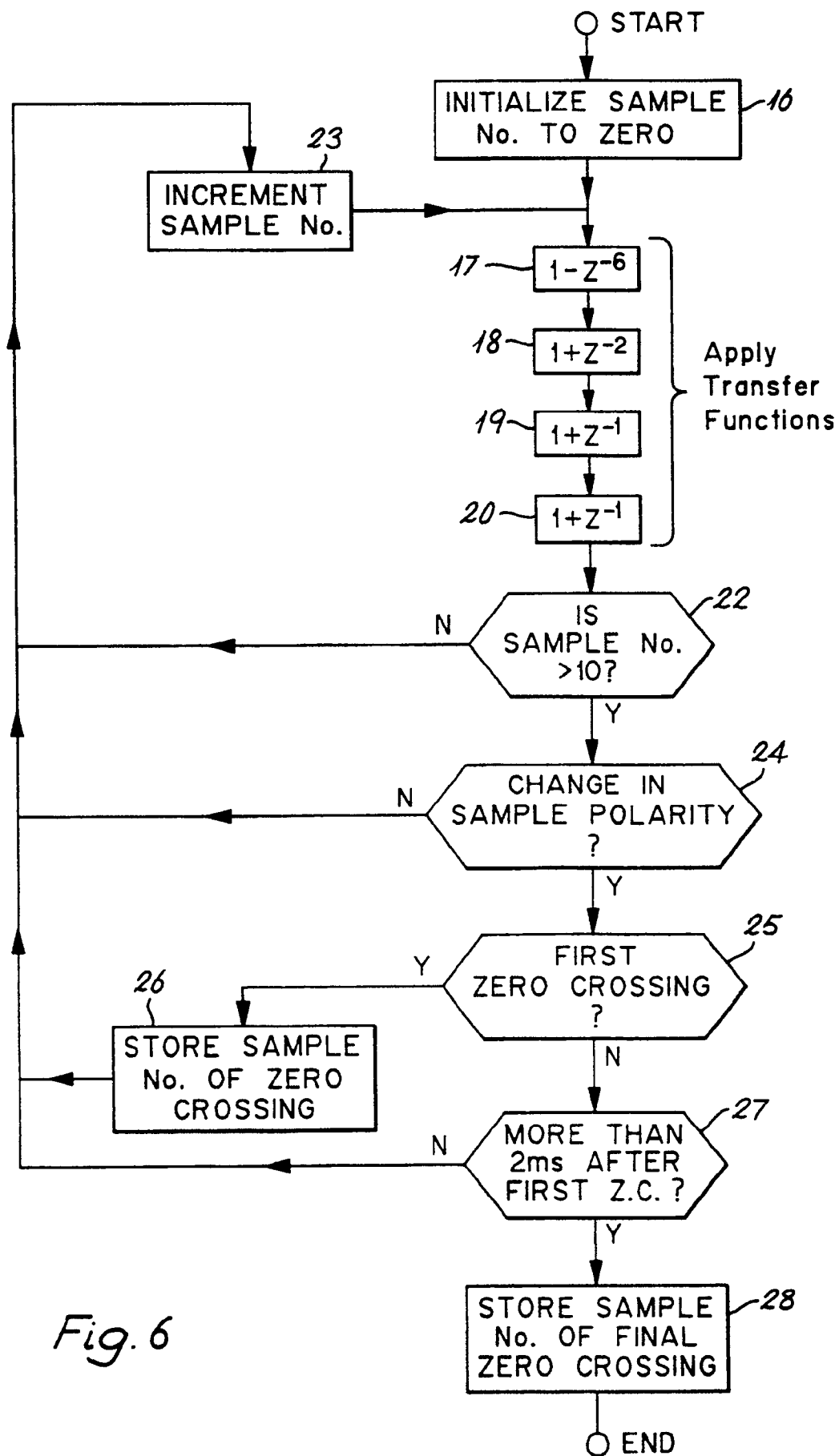
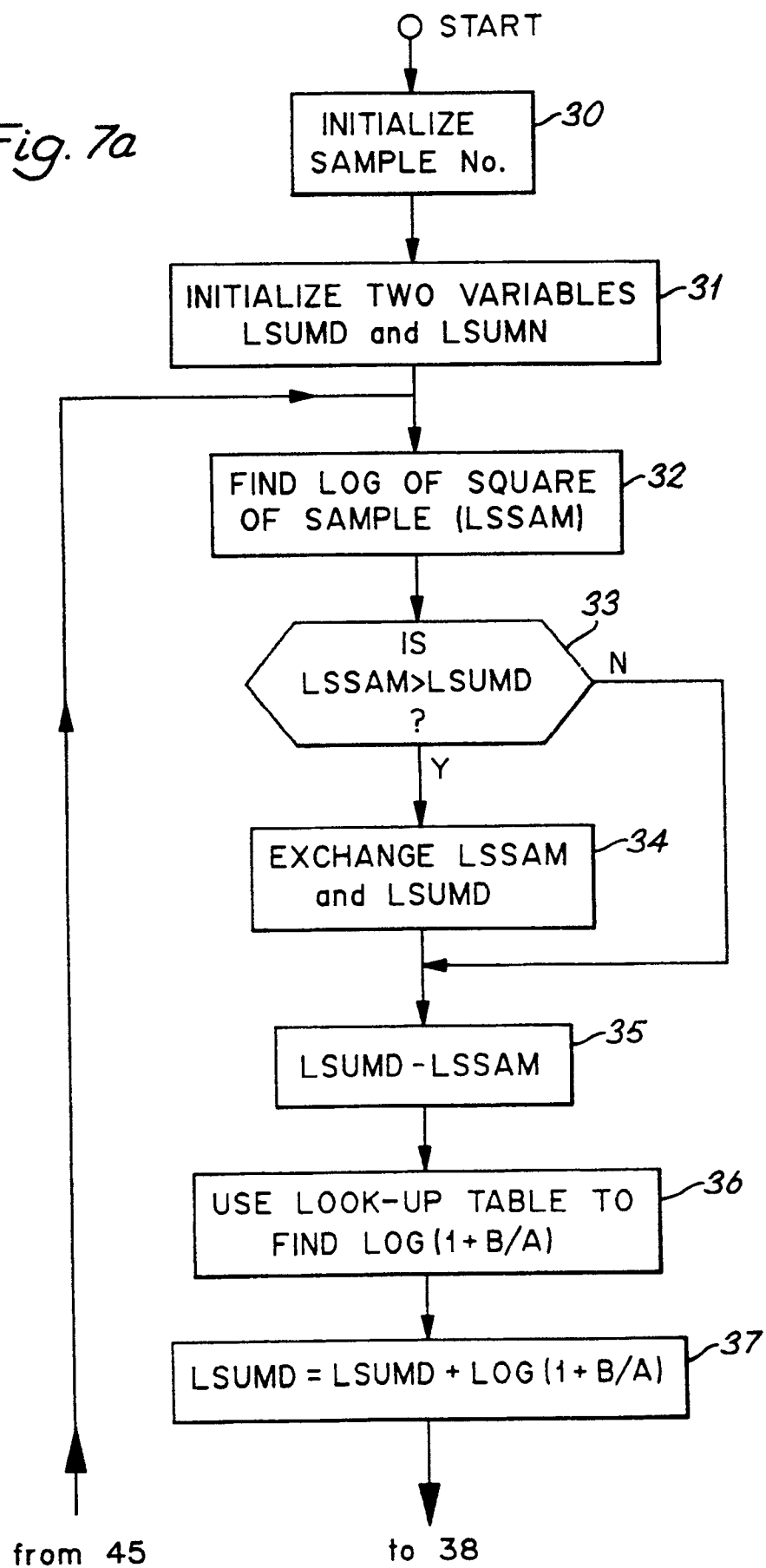


Fig. 6

Fig. 7a



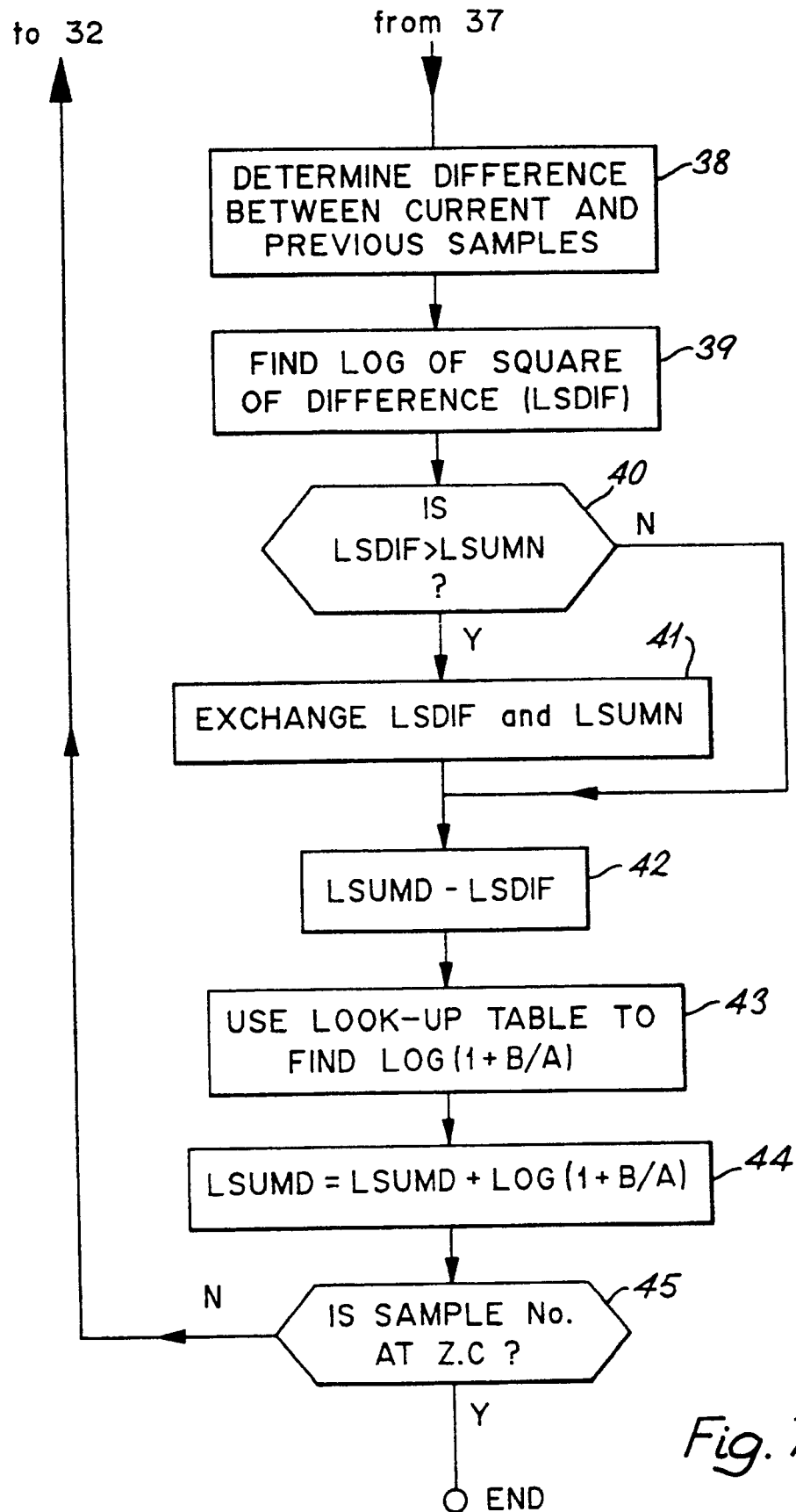


Fig. 7b

