

12 **EUROPEAN PATENT APPLICATION**

21 Application number: 86110212.7

51 Int. Cl.4: G10L 9/08

22 Date of filing: 24.07.86

30 Priority: 29.08.85 US 770530

43 Date of publication of application:
04.03.87 Bulletin 87/10

84 Designated Contracting States:
AT BE CH DE FR GB IT LI LU NL SE

71 Applicant: SCOTT INSTRUMENTS
CORPORATION
1111 Willow Springs Drive
Denton Texas 76201(US)

72 Inventor: Scott, Brian Lee
1500 Sandy Creek Denton
Denton County Texas 76205(US)
Inventor: Goodman, Robert Gray
1925 Parkside Drive Denton
Denton County Texas 76201(US)
Inventor: Newell, John Mark
609 Neptune Denton
Denton County Texas 76201(US)
Inventor: Smith, Lloyd Allen
615 Rose Denton
Denton County Texas 76201(US)

74 Representative: UEXKÜLL & STOLBERG
Patentanwälte
Beselerstrasse 4
D-2000 Hamburg 52(DE)

54 **Method and apparatus for generating a signal transformation and the use thereof in signal processings.**

57 A method and apparatus for generating a signal transformation useful in signal processing is provided. According to the preferred embodiment, a signal, e.g., a speech waveform, is first converted into a sequence of digital data samples, and a reference position along a first sub-part of the sequence is then selected. A "weighted" histogram corresponding to the reference position is then generated according to a correlation function. Thereafter, a new reference position is selected, for example, at a sub-part of the sequence located a pitch period of the signal from the original reference position, and an additional histogram is generated for this sub-part. The plurality of histograms comprise the transformation of the signal, which retains a substantial part of the informational content of the original signal. Therefore, the transformation is then used as the

signal itself in signal processing applications such as speech compression and synthesis.

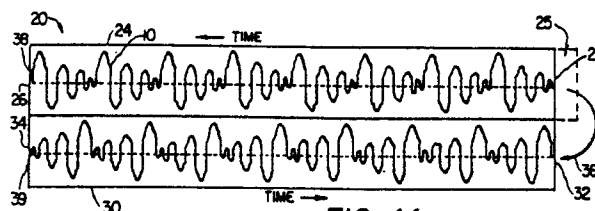


FIG. 1A

METHOD AND APPARATUS FOR GENERATING A SIGNAL TRANSFORMATION AND THE USE THEREOF IN SIGNAL PROCESSING

TECHNICAL FIELD

The present invention relates to signal processing techniques and particularly to a method and apparatus for generating a signal transformation which retains a substantial part of the informational content of the original signal.

BACKGROUND OF THE INVENTION

Audition is a temporally-based sense, whereas vision is primarily spatially-based. In perceiving speech, temporal events as brief as a few thousandths of a second are critical for making simple phonetic or word-based distinctions, such as between "pole" and "bowl," or "tow down" and "towed down." In addition to its highly developed temporal-resolving power, the ear also exhibits excellent spectral resolution and dynamic range. Exactly how the ear exhibits such fine spectral resolution without sacrificing temporal resolution remains a mystery. If more were understood about how the ear works, such knowledge could be applied to speech technologies to improve the performance of speech recognizers and coding devices.

Satisfactory temporal information from an acoustic speech signal is important for performing certain types of speech processing, e.g., speech segmentation in phonetically-based recognition systems. Likewise, satisfactory spectral resolution of the speech signal is important for other types of speech processing such as speech compression and synthesis. Current state-of-the-art digital signal processors cannot support such diverse speech processing applications because all suffer the classical trade-off of frequency versus time resolution --processors exhibiting good frequency resolution have poor temporal resolution, and vice versa. A digital signal processor having good spectral and temporal resolution would be a tremendous benefit to the speech industry because it would allow a single processing system to approximate the performance characteristics of the ear itself.

An ideal digital signal processor for use in speech processing would provide a unique representation or "transformation" of the speech signal from which all relevant speech features could be

derived. As is well known in the art, these features include voice pitch, amplitude envelope, spectrum and degree of voicing. It is presently common in speech systems to use totally different representations of the speech signal to abstract these features, depending on the type of speech processing application being implemented, and the capabilities of the processor carrying out the implementation.

There is therefore a need for a method and apparatus for generating a speech signal transformation which retains a substantial part of the informational content of the original signal, thereby facilitating extraction, from the transformation itself, of the speech features required for varied speech processing applications such as compression and synthesis.

BRIEF SUMMARY OF THE INVENTION

According to the present invention, a method and apparatus is provided for generating a signal transformation which retains a substantial part of the informational content of the original signal required for speech processing applications. As used herein, such applications include speech compression, speech synthesis and speech segmentation. In the preferred embodiment, the transformation is generated by converting all or part of the original signal into a sequence of data samples, selecting a reference position along a first sub-part of the sequence, and generating a histogram for the reference position according to a correlation function. Thereafter, a reference position along a second sub-part of the sequence is selected, and an additional histogram is generated for this reference position. The plurality of histograms generated in this fashion comprise the transformation. According to the invention, the transformation is then used as the signal itself in signal processing applications.

In one embodiment, the transformation comprises a plurality of "weighted" histograms, each having a predetermined number of positions d_{\max} and being derived from a general class of "differencing" functions of the form:

$$\text{histogram}(d,a) = \frac{1}{\text{scnt}} \sum_{n=n_0 - \left\lfloor \frac{\text{scnt} + a \cdot d}{2} \right\rfloor}^{n_0 + \left\lfloor \frac{\text{scnt} + a \cdot d}{2} \right\rfloor} |x(n) - x(n+d)| \quad (1)$$

for $d = 1, 2, \dots, d_{\max}$

where: $x(\cdot)$ is the sampled signal,
 n_0 is a reference position,
 d = time lag between samples,
 scnt = averaging interval in samples, and
 $a = 0, 1$.

The present invention also includes suitable apparatus for deriving "weighted" histograms according to expression (1) above. In a preferred embodiment, the data samples representing a first sub-part of the sequence are applied sequentially through a differencing correlator having first and second sections, the output of the first section connected to the input of the second section through a temporary storage area. A new data sample is then applied to the first correlator section and the remaining samples therein shift by one position. A data sample is thereby removed from the first correlator section to the temporary storage area for a first iteration of the differencing calculation. The magnitudes of the data samples in the second correlator section are then differenced with the magnitudes of positionally-corresponding data samples in the first correlator section, and absolute values of these differences are then calculated to produce "even" values which are then added to the histogram for the reference position. Thereafter, the data sample in the temporary storage area (for the first iteration) is applied to the second correlator section and the remaining samples therein shifted by one position. The "differencing," "absolute value" and "summation" steps are then repeated to produce "odd" values of the histogram. This operation (i.e., summation to the "even" and "odd" values) represents one complete cycle of the histogram, and is repeated "scnt" times according to expression (1) to complete the formation of the histogram for the reference position along the first sub-part of the data sample sequence. The process is then repeated for reference positions along other sub-parts of the sequence, each reference position preferably located a pitch period (or multiple thereof) apart, to form additional histograms.

Referring back to equation (1), when $a=0$, the function "histogram (d,a)" reduces to the well-known average magnitude difference function - (AMDF). When $a=1$, the function "histogram (d,a)" produces a so-called sliding average magnitude

difference function (SAMDF), which differs from the AMDF in that the center point of the samples used to compute "histogram (d,1)" is the same for all values of "d." This center point is preferably the reference position, or " n_0 " in expression (1).

According to an important feature of the present invention, the plurality of "weighted" histograms comprise the transformation of the original signal. It has been found that transformations of the type disclosed herein retain a substantial part of the informational content of the original signal, with only the phase information removed. The transformation is then used according to the invention by various speech or other signal processing applications. For example, to form a compressed version of the original signal, a predetermined portion of each histogram generated every other pitch period of the signal is then stored. Conversely, to implement speech synthesis, the compressed transformation is reconstructed. In neither case, however, does the method require costly and complex conversion of the signal between the time and frequency domains, as in the prior art.

According to the invention a special purpose microprocessor is also provided which, under the control of a software routine, generates the histograms. A general purpose microprocessor is also provided for effecting overall system control, and for controlling specialized processing applications, such as signal compression and synthesis. These microprocessors operate concurrently in a full duplex digital transceiver configuration to facilitate real-time communications to and from the system.

BRIEF DESCRIPTION OF THE DRAWINGS

For a more complete understanding of the present invention and the advantages thereof, reference is now made to the following Description taken in conjunction with the accompanying Drawings in which:

FIGURE 1A discloses a correlator structure of the present invention having first and second sections for use in generating histograms according to the present invention.

FIGURE 1B is a histogram, partially cut-away, generated by the correlator of FIGURE 1A.

FIGURE 2 is a flowchart diagram detailing the steps used to calculate the sliding average magnitude difference function (SAMDF) according to the present invention.

FIGURE 3 is a block diagram of a speech system of the present invention for performing speech processing applications such as compression and synthesis.

FIGURE 4 is a flowchart diagram of a signal compression routine which uses the signal transformation generated by the SAMDF process to compress the original signal waveform.

FIGURE 5 is a flowchart diagram of a signal synthesis routine for synthesizing the signal compressed by the signal compression routine diagrammed in FIGURE 4.

DETAILED DESCRIPTION

Referring now to the drawings wherein like reference characters designate like or similar parts throughout the several views, FIGURE 1A discloses a correlator structure for generating histograms according to the present invention. As will be described, a plurality of such histograms form a so-called "transformation" of the signal which retains a substantial part of the informational content thereof. For the purpose of explanation only, and not by way of limitation, the technique is explained below with an emphasis on human speech as the source waveform. It should be appreciated, however, that the method and apparatus of the present invention is fully applicable to all types of analog and digital source signals, regardless of how such signals are derived.

In the preferred embodiment, histograms are generated according to one of a plurality of correlation functions. One subset of these functions are so-called "differencing" functions which operate to produce weighted histograms, having d_{\max} positions, of the form:

$$\text{histogram}(d, a) = \frac{1}{\text{scnt}} \sum_{n=n_0 - \left\lfloor \frac{\text{scnt} + a \cdot d}{2} \right\rfloor}^{n_0 + \left\lfloor \frac{\text{scnt} + a \cdot d}{2} \right\rfloor} |x(n) - x(n+d)| \quad (1)$$

for $d = 1, 2, \dots, d_{\max}$

where: $x(\cdot)$ is the sample signal,
 n_0 is a reference position,
 d = time lag between samples,
 scnt = averaging interval in samples, and
 $a = 0, 1$

According to an important feature of the present invention, it has been found that when a signal is processed to produce "weighted" histograms according to one of a plurality of correlation functions, such as the "differencing" function(s) of expression (1), the resulting transformation (which comprises a plurality of such histograms) retains a substantial part of the informational content of the original signal, with only the phase information removed. According to the invention, the transformation is then used as the original signal itself, thus

obviating costly and complicated conversion of the signal, or conversion of features extracted therefrom, between the time and frequency domains prior to and/or following processing.

Although the subsequent discussion is directed to apparatus for implementing the differencing function(s) defined in expression (1), the present invention envisions that histograms comprising the signal transformation are generated by different types of known "auto" or "cross" correlation functions of the general form:

$$\text{histogram}(d) = \lim_{N \rightarrow \infty} \frac{1}{2N+1}$$

$$u(n)v(n+d) \quad (2)$$

If "u" is identical to "v" in expression (2), "histogram(d)" reduces to the well-known auto-correlation function. If "u" is not identical to "v", expression (2) represents a cross-correlation function.

Referring now to FIGURE 1A, a schematic diagram is shown of a correlator 20 for use in the present invention for generating histograms according to the "differencing" function of expression (1) when $a=1$. The correlator 20 includes a first section 24 having a top entrance 26 and a top exit 28. The correlator 20 also includes a second section 30 having a bottom entrance 32 and bottom exit 34. As designated by the arrow 36, the top exit 28 of the first correlator section 24 is connected to the bottom entrance 32 of the second correlator section 30. As also shown in FIGURE 1A, the first correlator section 24 includes a temporary storage area 25 adjacent the exit 28 for temporarily storing a data sample, for the reasons to be described below.

For purposes of explanation only, the speech waveform 10 is shown in analog form inside the correlator 20. It should be appreciated, however, that in the actual method and apparatus of the present invention, the speech waveform 10 is first converted into a sequence of digital data samples. As seen in FIGURE 1A, a sub-part of the speech waveform 10 is passed sequentially through the first correlator section 24, through the temporary storage area 25, and then into the second correlator section 30. As each new data sample enters the top entrance 26 of the first section 24, the remaining data samples in this correlator section are each shifted one position towards the exit 28. A data sample is then removed to the temporary storage area 25 and held there for a predetermined time period to be described. According to a feature of the present invention, data samples in the second correlator section 30 are then differenced with positionally-corresponding data samples in the first correlator section 24. As used herein, the term "positionally-corresponding" refers to data samples in the respective correlator sections at any moment in time located the same distance from the ends of the correlator. Therefore, the data sample 38 located adjacent the top entrance 26 of the first section 24 "positionally-corresponds" to the data sample 39 located adjacent the bottom exit 34 of the second section 30.

Referring briefly now to FIGURE 1B, correlation of the speech waveform in the first and second correlator sections 24 and 30 produces a histogram 40 having a plurality of predetermined "buckets" or positions from $d=1,2,\dots$ to d_{\max} . The positions " d_1, d_3, d_5, \dots " represent the "odd" values of the histogram with the positions " d_2, d_4, d_6, \dots " representing the even values thereof. Although not shown in detail in FIGURE 1B, the length of the histogram 40 is normally two times the length of each correlator section. Also, the length of each sub-part of the data sample sequence is typically greater than " d_{\max} ".

With respect to expression (1), if $a=0$, the function reduces to the well-known average magnitude difference function (AMDF). If $a=1$, the function reduces to a so-called sliding average magnitude difference function (SAMDF), which differs from the AMDF in that the center point of the samples used to compute "histogram(d,1)" is the same for all values of "d". Because of this common reference, the SAMDF is used in the preferred embodiment of the invention and now described with respect to FIGURE 2.

The SAMDF scheme begins at step 41 - (assuming the correlator is filled with a portion of a first sub-part of the sequence) by initializing the d_{\max} positions of the histogram to zero. In step 42, a new data sample is moved into the first correlator section 24 and the remaining samples therein are shifted by one position. Step 42 therefore moves a data sample into the temporary storage area 25 for the first iteration of the calculation. In step 43, differences in magnitude between corresponding samples in the correlation sections are calculated. In particular, the magnitude of the first sample in the correlator section 24 adjacent the top entrance 26 thereof is differenced from the magnitude of the last sample in the correlator section 30 adjacent the bottom exit 34 thereof. This differencing step is also carried out for the rest of the samples at each position in the correlator sections. In step 44, the absolute values of the differences calculated in step 43 for each position in the correlator are then determined and in step 45, added to the summation to produce the "even" positions " d_2, d_4, d_6, \dots " of the histogram 40. Thereafter, an inquiry 46 is made to determine if a complete cycle of the histogram formation has been run. If not, the routine branches back to step 47, where the data sample in the temporary storage area 25 (received during the first iteration) is shifted into the second correlator sec-

tion 30 and the remaining samples therein shifted by one position. Steps 43-45 are then repeated to increment the "odd" values "d₁, d₃, d₅..." of the histogram 40. If the result of inquiry 46 is positive, a test 48 is then made to see if "scnt" samples have been applied to the temporary storage area 25; if not, the routine branches back to step 42, and the method repeats as described above. If the result of inquiry 48 is positive, the histogram may be normalized (for example, by dividing each histogram value by "scnt") to produce the completed histogram for the first sub-part of the data sample sequence originally applied through the correlator sections. This process is then repeated in step 49 for additional sub-parts of the signal (applied through the correlator sections) to produce additional histograms comprising the signal transformation.

In the preferred embodiment, reference positions along the sample sequence are separated by a pitch period, or multiple thereof, of the signal. Also, when the SAMDF process of FIGURE 2 is implemented, the data sample moved into the temporary storage area 25 after "scnt/2" cycles represents the reference position along the sub-part of the sequence.

Referring briefly back to expression (I), it should be appreciated that there are other methods for implementing this expression besides the method steps shown in FIGURE 2. For example, rather than shifting data samples into a correlator structure and producing the various summations as described, the expression may be calculated by initializing the histogram to its first position "d₁" (i.e., setting d=1), and summing over the range of "n" as shown in expression (I). This produces the value "histogram (I,a)". Thereafter, the histogram can be initialized to its second position "d₂" (d=2), and the process repeated until calculation of the histogram is completed.

Referring now to FIGURE 3, a schematic block diagram is shown of a speech system 50 designed to provide the capabilities needed to produce the signal transformation according to the present invention, and also to provide the capabilities needed for using this transformation in speech processing applications. As discussed above, for the purposes of explanation only system 50 will be described in the context of a speech development system. System 50, however, is fully capable of interfacing with all types of signal processing applications, and the reference to speech-related applications herein is not meant to be limiting.

The speech system 50 includes a general purpose microprocessor 52 which has several input/output (I/O) devices tied thereto. Speech system 50 includes a pair of serial digital communication links 54 and 56 connected to the general

purpose microprocessor 52 through universal asynchronous receiver/transmitters (UART's) 58 and 60, respectively. Such devices are well known and serve to interface the parallel word-based microprocessor 52 to the serial bit communication links 54 and 56. Speech system 50 also includes an analog input path 62 to the general purpose microprocessor 52 comprising bandpass filter 64 and analog-to-digital (A/D) converter 66. An analog output path 68 is also provided from the general purpose microprocessor 52 comprising low pass filter 70 and digital-to-analog (D/A) converter 72. An analog speech waveform is applied to the analog input path 62, where it is band limited by the filter 64, and digitized by the A/D converter 66. The digitized version of the speech waveform may then be transmitted over one of the digital serial communication links 54 or 56 to a remote system similar to the speech development system 50.

As also seen in FIGURE 3, the general purpose microprocessor 52 includes an associated random access memory (RAM) 51 for storing application programs and data, and also a read only memory (ROM) 53 for storing operating programs which control the microprocessor 52.

According to a feature of the present invention, the speech system 50 includes a special purpose microprocessor 74 which, under the control of a software routine, carries out the SAMDF process of FIGURE 2. Special purpose microprocessor 74 includes an associated control store 76 for storing this routine, and an associated random access memory (RAM) 78 for communicating with the general purpose microprocessor 52. General purpose microprocessor 52 passes digital data samples from the analog input path 62 into the RAM 78 and these samples are then processed in the special purpose microprocessor 74 under the control of a routine stored in control store 76. The resulting transformation of the speech waveform is then stored back in the RAM 78. The contents of RAM 78 are then read by general purpose microprocessor 52 without interrupting the continued processing of additional portions of the waveform by special purpose microprocessor 74.

Accordingly, special purpose microprocessor 74 operates concurrently with general purpose microprocessor 52 to enable the microprocessor 74 to carry out the SAMDF correlation calculations while the microprocessor 52 provides other system control functions.

Speech system 50 provides full duplex digital transceiver operation for facilitating real-time communications to and from the system. When the system 50 is initialized, control programs are down loaded into the RAM 51 associated with the general purpose microprocessor 52. These programs control the microprocessor 52 to down load the

SAMDF routine into the control store 76 associated with special purpose microprocessor 74. The speech waveform is then received over the analog input path 62 and processed as described above.

According to an important feature of the present invention, once the signal transformation has been generated as discussed above, this transformation is then used as the signal itself by speech processing applications such as compression, synthesis and segmentation.

Referring now to FIGURE 4, a flowchart diagram is shown of a signal compression routine of the present invention which operates on the signal transformation to produce a compressed version of the original speech signal. As is known in the art, the object of speech compression is to represent analog speech with as few digital bits as possible. Prior art techniques, such as linear predictive coding (LPC), are based on the successful extraction of voice parameters from the speech signal and accurate voiced/unvoiced decisions. Although LPC and other prior art formant coding techniques provide effective speech signal compression in some applications, such techniques break down in noisy environments and when the speech signal is sampled at low data rates.

To ameliorate these and other problems of the prior art, the compression technique of the present invention takes advantage of certain informational redundancies inherent in the signal, which are also present in the signal transformation generated by the SAMDF process.

It has been found that a first source of informational redundancy in a speech signal exists because the speech waveform is substantially similar in any two contiguous pitch periods. Therefore, the storing of every other pitch period of the speech waveform represents a way to compress speech by a factor of 2:1. A second source of informational redundancy in the speech waveform is based on the notion that speech is normally a bipolar, approximately symmetrical waveform about an arbitrary reference level. If the waveform is rectified and zeros are eliminated therefrom, then the original waveform can be compressed by another factor of two, or by a total factor of 4:1. A third source of informational redundancy within the speech waveform is inherent in the way voiced signals are produced by the larynx. The glottal source has two phases, an open phase and a closed phase, and the resonances of the vocal tract are best represented in the speech waveform while the glottis is closed. Therefore, because the glottis is closed roughly 50% of the pitch period, only half of the speech waveform is carrying information during the

pitch period itself. Accordingly, the storage of only one-half of a pitch period represents a way to compress the speech waveform by another factor of two, for a total compression ratio of 8:1.

Referring back to FIGURE 1B, the SAMDF process correlates positive and negative phases of an input speech waveform, resulting in the histogram 40 with minimas corresponding to half cycles from the waveform. Accordingly, use of the SAMDF correlation process exploits the positive-to-negative cycle redundancy inherent in the speech waveform. Moreover, as also seen in FIGURE 1B, the SAMDF process produces a highly symmetrical histogram 40, such that storage of only one-half of a pitch period represented in the histogram is required. Storage of one-half of a pitch period thus exploits the redundancy in the waveform resulting from the physical characteristics of the glottal source. Further, in the preferred embodiment of the invention, the histogram 40 is generated by the correlator 20 by selecting reference positions along the data sample sequence every other pitch period, such that the histogram represents an "averaged" correlation over two pitch periods. This feature of the invention thus exploits the pitch period-to-pitch period redundancy inherent in the input speech waveform resulting in a total compression ratio of 8:1.

The compression routine in FIGURE 4 begins at instruction 80 wherein data samples are moved into the RAM 78, where they are processed by the special purpose microprocessor 74. As discussed above with respect to FIGURE 3, the data samples are obtained from conversion of an analog sound wave by the A/D converter 66. The SAMDF correlation is then carried out in step 80 by the special purpose microprocessor 74 of FIGURE 3 under the control of a software routine stored in the associated control store 76.

After each new data sample is moved into the RAM 78, a check 84 is made to determine whether or not a completed histogram (as described with respect to FIGURE 2) is ready for further processing. If the histogram is not ready, control returns to step 80, and another data sample is moved into the RAM 78 as previously described by step 42 in FIGURE 2. When the histogram is ready, i.e., the test in step 84 is positive, the histogram is moved from the RAM 78 to the RAM 51 in step 88, so that it can be processed by the general purpose microprocessor 52.

Referring back to FIGURE 4, the signal compression routine continues in step 90 to determine whether it is time to track the pitch of the waveform. If the result of the inquiry 90 is negative, i.e., if the time interval for tracking pitch has not

elapsed, the routine branches to step 92 wherein one-half of the pitch period is encoded from the histogram, preferably by using two-bit adaptive differential pulse code modulation (ADPCM).

Encoding of the compressed waveform incurs some overhead; for example, the frequency, or length of the pitch period, must be stored with the encoded waveform. In order to minimize this overhead, and because the pitch of voiced speech does not change rapidly, the system preferably tracks the pitch of the input speech signal only at certain time intervals, which may vary from as frequently as each pitch period to as infrequently as several pitch periods.

Returning to FIGURE 4, if the result of inquiry 90 is positive, then the routine continues with step 94 to determine the pitch period. In step 96, the routine continues by feeding the pitch period determined in step 94 back to the special purpose microprocessor 74. In step 98, the pitch is encoded with the routine continuing in step 100 to calculate the maximum amplitude in the pitch period, or gain factor. In step 102, the gain factor is then encoded, preferably using a log(base 2) representation, and the routine continues with step 92 as discussed above. Following step 92, an inquiry 104 is made to determine whether compression is complete. If not, the routine recycles back to step 80 wherein additional portions of the speech signal are digitized and the compression routine continues as described above. If compression is completed, then the routine terminates at step 106.

As detailed in the flowchart diagram of FIGURE 4, the first analysis performed on the histogram is pitch extraction. Pitch is determined by examining minimas in the histogram, analyzing for harmonic relations and selecting a first pitch trough. This value is then used to control the amount of time over which the next histogram will be summed. An effect of the process is to produce highly symmetrical histograms, so that only one-half of the pitch period in the histogram need be stored. This provides a 2:1 factor of compression in the speech waveform. Moreover, according to the method, histograms are output every other pitch period to provide another 2:1 factor of compression, or a total compression ratio of 4:1. As also noted above, the encoding step 92 codes the histograms using a two-bit ADPCM scheme modulation scheme. This represents another factor of four compression on the original eight-bit digitized waveform. Thus, the total compression ratio of the technique is 16:1.

Referring now to FIGURE 5, a flowchart diagram of a signal synthesis routine of the present invention is shown. As discussed above, this routine operates on the SAMDF signal transformation generated by the special purpose microprocessor, and in particular on the transformation as com-

pressed by the compression routine set forth in FIGURE 4. Synthesis begins with instruction 110, wherein the routine is initialized by receiving data representing the compressed speech signal. The routine continues with inquiry 112 which determines whether the pitch period should be read. If the result of the inquiry 112 is positive, the routine continues in step 114 to read the pitch period from the bitstream data received over one of digital serial communication links of FIGURE 3. Thereafter, the gain factor is read in step 116 from the bitstream data. Following step 116, or if the result of inquiry 112 is negative, the method continues in step 118, wherein one-half of the pitch period for the compressed segment is expanded from the bitstream data. In step 120, the pitch of the segment is interpolated, as is the gain factor in step 122. The routine continues in step 124 to synthesize the pitch period(s). Following step 124, the routine enters inquiry 126 to determine whether the speech waveform synthesis has been completed. If not, the method returns to step 110 to get data to synthesize the next segment. If the synthesis is complete, the routine terminates at step 128.

Accordingly, synthesis occurs in four steps. Preferably, the stored encoded pitch and gain factors are first read and decoded. The second step consists of a simple expansion of the histogram from ADPCM to pulses code modulation (PCM) format, which is accomplished in step 118 of FIGURE 5. Thereafter, the reconstructed waveform is reflected in step 124 to form the pitch period. The fourth and final step is to repeat the pitch period, with the process then repeated for each subsequent portion of the compressed speech waveform.

Accordingly, the present invention provides a method and apparatus for generating a transformation of a signal waveform useful in speech processing for example, compression and synthesis. This transformation retains the informational content of the original signal and therefore is used directly to represent the signal. The "use" of the signal transformation as the signal itself obviates costly and complex computational algorithms for converting the signal (or features thereof) between the time and frequency domains prior to and following the signal processing application(s). In the preferred embodiment of the invention, a special purpose microprocessor is provided to run a software routine for generating the transformation by calculating a sliding average magnitude difference function - (SAMDF) histogram for continuous segments of the speech waveform.

As discussed above, although the method and apparatus of the present invention has been described in detail with respect to speech processing applications such as compression/synthesis, it should be appreciated that the techniques described here-

in are fully compatible with all types of signal processing applications. Accordingly, the scope of the present invention is not limited to use of the signal transformation to effect speech compression/synthesis.

Although the invention has been described and illustrated in detail, it is clearly understood that the same is by way of illustration and example only and is not to be taken by way of limitation. The spirit and scope of the present invention are to be limited only by the terms of the appended claims.

Claims

1. A method for generating a transformation of a signal and using the transformation in signal processing, comprising the steps of:

- (a) converting a portion of the signal into a sequence of data samples;
- (b) selecting a reference position along a first sub-part of said sequence;
- (c) generating a histogram for the reference position according to a correlation function;
- (d) selecting a reference position along a second sub-part of said sequence;
- (e) repeating step (c) for the reference position along the second sub-part of the sequence to produce an additional histogram, whereby all of the

histograms comprise the transformation which retains a substantial part of the informational content of the signal; and

(f) processing the transformation as the signal itself in signal processing application.

2. The method for generating a transformation of a signal as described in Claim 1 wherein the correlation function is an average magnitude difference function (AMDF).

3. The method for generating a transformation of a signal as described in Claim 1 wherein the correlation function is a sliding average magnitude difference function (SAMDF).

4. The method for generating a transformation of a signal as described in Claim 1 wherein the correlation function is an auto-correlation function.

5. A method for generating a transformation of a signal and using the transformation in signal processing, comprising the steps of:

- (a) converting a portion of the signal into a sequence of digital data samples;
- (b) selecting a reference position along a first sub-part of said sequence;
- (c) generating a histogram, having d_{\max} positions, for the reference position according to the expression:

$$\text{histogram}(d, a) = \frac{1}{\text{scnt}} \sum_{n=n_0 - \left\lfloor \frac{\text{scnt} + a \cdot d}{2} \right\rfloor}^{n_0 + \left\lfloor \frac{\text{scnt} + a \cdot d}{2} \right\rfloor} |x(n) - x(n+d)| \quad (1)$$

for $d = 1, 2, \dots, d_{\max}$

where: $x(\cdot)$ is the sampled signal,
 n_0 is a reference position,
 d = time lag between samples,
 scnt = averaging interval in samples, and
 $a = 0, 1$

(d) selecting a reference position along a second sub-part of said sequence;

(e) repeating step (c) for the reference position along the second sub-part of the sequence to produce an additional histogram, whereby all of the histograms comprise the transformation which retains a substantial part of the information content of the signal; and

(f) processing the transformation as the signal itself in signal processing application.

6. The method for generating a transformation of a signal as described in Claim 5 wherein said step of generating a histogram operates over "scnt" cycles for each reference position.

7. The method for generating a transformation of a signal as described in Claim 6 wherein said step of generating a histogram for each reference position includes the steps of:

(g) applying the digital data samples through a correlator having first and second sections and a temporary storage area, an output of said first correlator section forming an input to said second correlator section through the temporary storage area;

(h) moving a new data sample into the first correlator section and shifting the remaining data samples therein by one position, whereby a data sample is removed from said first correlator section to said temporary storage area.

8. The method for generating a transformation of a signal as described in Claim 7 wherein each said cycle of the histogram includes the steps of:

(i) differencing a magnitude of each data sample in said second correlator section with a magnitude of a positionally-corresponding data sample in said first correlator section;

(j) determining an absolute value of each difference calculated in step (i) to produce "even" values of the histogram;

(k) moving said data sample in said temporary storage area into the second correlator section and shifting the remaining data samples therein by one position;

(l) repeating steps (i) -(j) to produce "odd" values of the histogram.

9. The method for generating a transformation of a signal as described in Claim 5 wherein said signal processing application include the step of compressing the signal.

10. The method for generating a transformation of a signal as described in Claim 9 wherein said step of compressing includes the step of storing a predetermined portion of each histogram generated in step (c) as a compressed version of said speech signal.

11. The method for generating a transformation of a signal as described in Claim 10 wherein said step of compressing includes selecting the reference position along said sequence every other pitch period of said signal.

12. The method for generating a transformation of a signal as described in Claim 10 wherein said signal processing applications include the step of synthesizing said transformation from said compressed version of said speech signal.

13. Apparatus for generating a transformation of a signal and using the transformation in signal processing, comprising:

means for converting said signal into a sequence of digital data samples;

means for selecting a reference position along a first sub-part of said sequence;

means for generating a histogram for said reference position according to a correlation function;

means responsive to said generating means for selecting a reference position along a second sub-part of said sequence and generating an additional histogram for the second sub-part, whereby said histograms comprise said transformation which retains a substantial part of the informational content of the signal; and

means for processing said transformation as said signal itself in signal processing applications.

14. The apparatus for generating a transformation of a signal as described in Claim 13 wherein said processing means includes:

means for producing a compressed version of the transformation to represent a compressed version of the signal.

15. The apparatus for generating a transformation of a signal as described in Claim 14 wherein said processing means further includes:

means for synthesizing the signal compressed by the means for producing a compressed version to resynthesize the signal transformation.

16. The apparatus for generating a transformation of a signal as described in Claim 13 wherein said means for generating includes a special purpose microprocessor, having an associated storage device, for implementing one of said plurality of correlation functions.

17. The apparatus for generating a transformation of a signal as described in Claim 16 wherein said means responsive to said means for generating includes a general purpose microprocessor having an associated storage device.

18. The apparatus for generating a transformation of a signal as described in Claim 17 wherein said general purpose microprocessor and said special purpose microprocessor operate concurrently to provide real-time generation of said signal transformation.

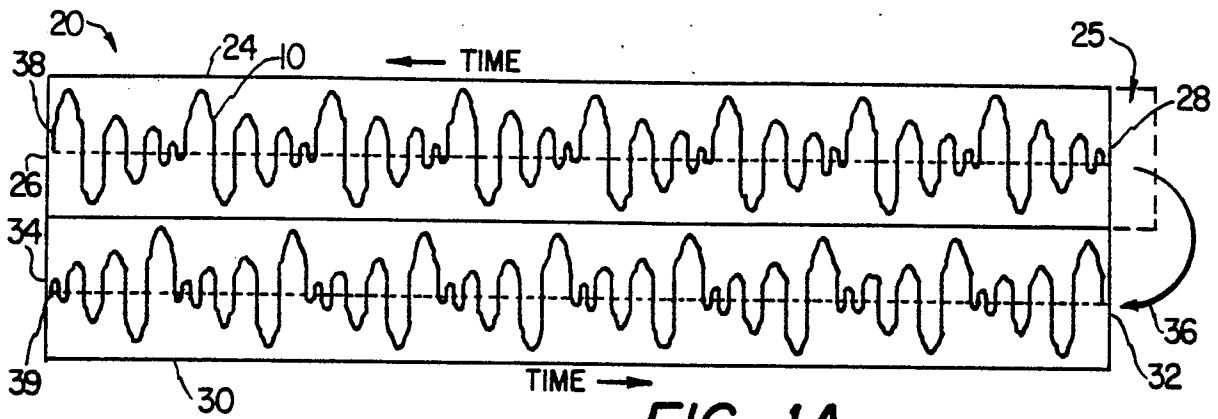


FIG. 1A

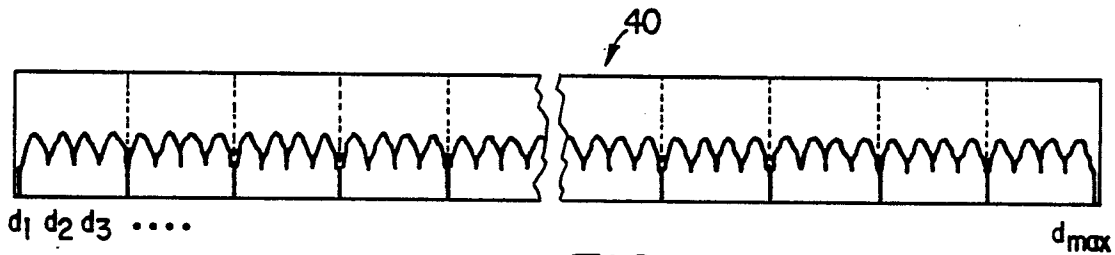


FIG. 1B

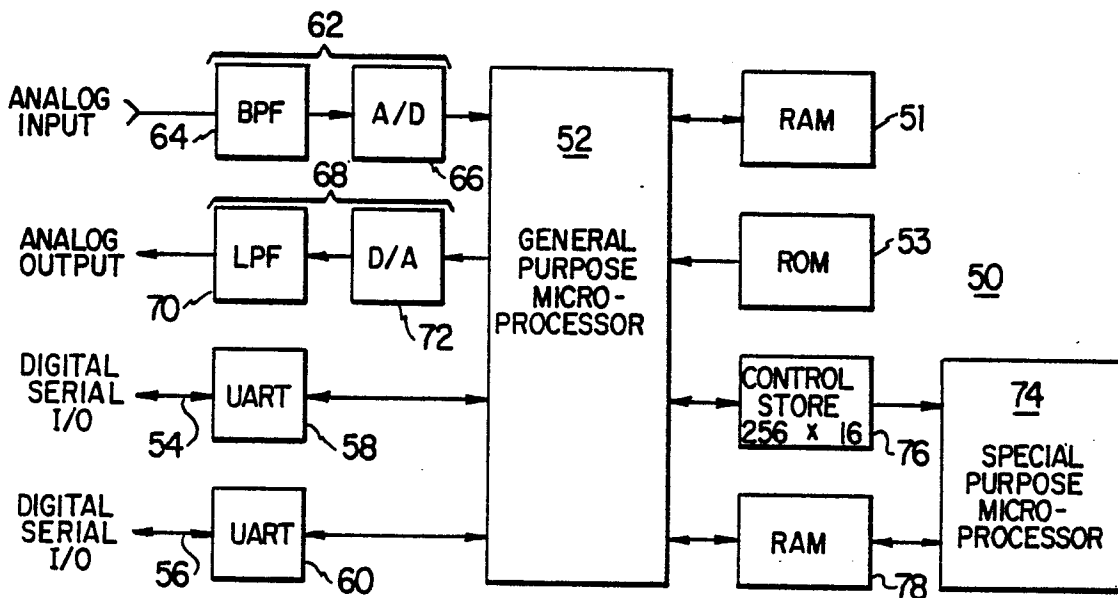


FIG. 3

