

Serial No. 553.161
Filing Date 7 November 1995
Inventor George S. Kang
Lawrence J. Fransen.

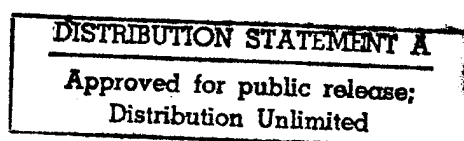
NOTICE

The above identified patent application is available for licensing. Requests for information should be addressed to:

OFFICE OF NAVAL RESEARCH
DEPARTMENT OF THE NAVY
CODE OCCC3
ARLINGTON VA 22217-5660

19960418 058

DTIC QUALITY INSPECTED 1



Serial No.:
Inventors: Kang and Fransen

PATENT APPLICATION
Navy Case No. 77,023

1 METHOD AND APPARATUS FOR SEGMENTING A SPEECH WAVEFORM

3 BACKGROUND OF THE INVENTION

4 1. Field of the Invention

5 The present invention is directed to a system for processing
6 human speech and, more particularly, to a system that pitch-
7 synchronously segments the human speech waveform into individual
8 pitch waveforms which may be transformed, replicated, and
9 concatenated to generate continuous speech with desired speech
10 characteristics.

12 2. Description of the Related Art

13 The ability to alter speech characteristics is important in
14 both military and civilian applications with the increased use of
15 synthesized speech in communication terminals, message devices,
16 virtual-reality environments, and training aids. Currently,
17 however, there is no known method capable of modifying utterance
18 rate, pitch period, or resonant frequencies of speech by
19 operating directly on the original speech waveform.

20 Typical speech analysis and synthesis are based on a model
21 that includes a vocal tract component consisting of an electrical
22 filter and a glottis component consisting of an excitation signal
23 which is usually an electrical signal generator feeding the
24 filter. A goal of these models is to convert the complex speech

Serial No.:
Inventors: Kang and Fransen

PATENT APPLICATION
Navy Case No. 77,023

1 waveform into a set of perceptually significant parameters. By
2 controlling these parameters, speech can be generated with these
3 models. To derive human speech model parameters accurately, both
4 the model input (turbulent air from the lungs) and the model
5 output (speech waveform) are required. In conventional speech
6 models, however, model parameters are derived using only the
7 model output because the model input is not accessible. As a
8 result, the estimated model parameters are not often accurate.

9 What is needed is a different way of representing speech
10 that does not represent speech as an electrical analog sound
11 production mechanism.

12 13 SUMMARY OF THE INVENTION

14 It is an object of the present invention to represent the
15 speech waveform directly by individual waveforms beginning and
16 ending with the pitch epoch. These waveforms will be referred to
17 as pitch waveforms.

18 It is another object of the present invention to segment the
19 speech waveform into pitch waveforms.

20 It is also an object of the present invention to perform
21 pitch synchronous segmentation to obtain pitch waveforms by
22 estimating the center of a pitch period by means of centroid
23 analysis.

24 It is an additional object of the present invention to use

Serial No.:
Inventors: Kang and Fransen

PATENT APPLICATION
Navy Case No. 77,023

1 the ability to segment the speech waveform to perform speech
2 analysis/synthesis, speech disguise or change, articulation
3 change, boosting or enhancement, timber change and pitch change.

4 It is an object of the present invention to utilize
5 segmented pitch waveforms to perform speech encoding, speech
6 recognition, speaker verification and text to speech.

7 It is a further object of the present invention to provide a
8 speech model that is not affected by pitch interference, that is,
9 segmented pitch waveform spectrum is free of pitch harmonics.

10 The above objects can be attained by a system that uses an
11 estimate of the pitch period and an estimation of the center of
12 the pitch waveform to segment the speech waveform into pitch
13 waveforms. The center of the pitch waveform is determined by
14 finding the centroid of the speech waveform for one pitch period.
15 The centroid is found by finding a local minimum in the centroid
16 histogram waveform, such that the local minimum corresponds to
17 the midpoint of the pitch waveform. The midpoint or center of
18 the pitch waveform along with the pitch period is used to segment
19 or divide the speech waveform. The speech waveform can then be
20 represented by a set of such pitch waveforms. The pitch waveform
21 can be modified by frequency enhancement/filtering, waveform
22 stretching/shrinking in speech synthesis. The utterance rate of
23 the speech can also be changed by increasing or decreasing the
24 number of pitch waveforms in the output.

Serial No.:
Inventors: Kang and Fransen

PATENT APPLICATION
Navy Case No. 77,023

BRIEF DESCRIPTION OF THE DRAWINGS

These and other objects, features and advantages of the invention, as well as the invention itself, will become better understood by reference to the following detailed description when considered in connection with the accompanying drawings wherein like reference numerals designate identical or corresponding parts throughout the several views and wherein:

Figure 1 depicts a speech waveform with delineated pitch waveforms and an associated pitch period;

Figures 2(a) and 2(b) respectively depict a low-pass filtered speech waveform and a centroid histogram waveform for the speech waveform;

Figure 3 depicts the typical hardware of the present invention in a preferred embodiment;

Figure 4 shows the pitch synchronous segmentation operation of the present invention performed by the computer system 40 of figure 3;

Figures 5(a), 5(b) and 5(c) illustrate utterance rate changes;

Figure 6 illustrates pitch waveform replication to generate continuous speech;

Figures 7(a), 7(b) and 7(c) depict pitch alteration;

Figure 8 depicts spectrum modifications;

Figure 9 shows the structural elements in the computer

Serial No.:
Inventors: Kang and Fransen

PATENT APPLICATION
Navy Case No. 77,023

1 system 40 of figure 3 for performing the operations of segmenting
2 and reconstructing a speech waveform;

3 Figure 10(a) illustrates timing circuits for generating
4 various timing signals used in the system of figure 11;

5 Figure 10(b) depicts control timing diagrams;

6 Figure 11 depicts a discrete component embodiment of the
7 invention;

8 Figure 12 depicts a first type of circuit for utilizing the
9 segmented pitch waveform samples of figure 11 to modify the
10 waveform spectrum;

11 Figure 13 depicts a second type of circuit for utilizing the
12 segmented pitch waveform samples of figure 11 to modify the
13 pitch;

14 Figure 14 depicts the components used for replicating and
15 concatenating pitch waveforms to generate continuous analog
16 speech;

17 Figure 15 illustrates an alternate approach to segmentation;
18 and

19 Figure 16 depicts different functions used in correlation.
20

21 DESCRIPTION OF THE PREFERRED EMBODIMENTS

22 This invention is directed toward a speech
23 analysis/synthesis model that characterizes the original speech
24 waveform. In this invention, the speech waveform is modeled as a

Serial No.:
Inventors: Kang and Fransen

PATENT APPLICATION
Navy Case No. 77,023

1 collection of disjoint waveforms, each representing a pitch
2 waveform. Note that a pitch period is the time duration or the
3 number of speech samples present in the pitch waveform. A
4 segmented waveform of one pitch period can represent neighboring
5 pitch waveforms because of the redundancy inherent in speech.
6 Speech is reconstructed by replicating within a frame and
7 concatenating from frame to frame the segmented pitch waveforms.

8 The present invention is a technique for segmenting pitch
9 waveforms from a speech waveform of a person based on pitch. The
10 inventors have recognized that the speech waveform, as
11 illustrated in figure 1, is a collection of disjoint waveforms
12 1-10 where waveforms 2-10 are the result of the glottis opening
13 and closing at the pitch rate. A purpose of the invention is to
14 segment individual pitch waveforms. As noted previously, a
15 segmented waveform of one pitch period T 123 can represent
16 neighboring pitch waveforms because of the slowly varying nature
17 of the speech waveform. One pitch waveform representing more
18 than one pitch waveform is an important aspect of speech
19 compression and synthesis. In the example of figure 1, eight
20 pitch waveforms 3-10 are substantially similar. Because one
21 pitch waveform represents speech that may have many pitch
22 waveforms, speech compression is possible. Modification of the
23 speech waveform can be accomplished by modifying only a portion
24 of the speech waveform or the representative pitch waveform, an

Serial No.:
Inventors: Kang and Fransen

PATENT APPLICATION
Navy Case No. 77,023

1 advantage in speech modification and synthesis.

2 The segmentation of a speech waveform into pitch waveforms
3 requires two computational steps: 1) the determination of the
4 pitch period; and 2) the determination of a starting point for
5 the pitch waveform. Determining the pitch period can be
6 performed using conventional techniques typically found in
7 devices called vocoders. To determine the starting point of the
8 pitch waveform, the center of the pitch waveform is determined
9 first, in accordance with the present invention, by centroid
10 histogram waveform analysis.

11 The location of the centroid (or center of gravity), as used
12 in mechanics, is expressed by centroid function
13

$$\eta = \frac{\int_{x_1}^{x_2} x f(x) dx}{\int_{x_1}^{x_2} f(x) dx} \quad \text{for } x_1 \leq x \leq x_2 \quad (1)$$

15 where $f(x)$ is a non-negative mass distribution, and $[x_1, x_2]$ is
16 the domain of variable x . In this invention, x is a time
17 variable, $f(x)$ is the speech waveform, and the domain $[x_1, x_2]$
18 where $x_2 - x_1$, is one pitch period. Since $f(x)$ cannot be negative,
19 a sufficient amount of bias is added to the speech waveform so

Serial No.:
Inventors: Kang and Fransen

PATENT APPLICATION
Navy Case No. 77,023

1 that $f(x) > 0$.

2 Figures 2(a) and 2(b) respectively show a low pass filtered
3 speech samples $S'(\cdot)$ 126 and centroid histogram samples $C(\cdot)$ 182
4 of the centroid function. The center of a pitch period is
5 defined to be at a local minimum of the centroid histogram
6 samples $C(\cdot)$ 182 of the speech waveform for that pitch period. A
7 local minimum location α 199 of the samples $C(\cdot)$ 182 occurs at a
8 midpoint 20 of a pitch period T 123. Knowing the location α 199
9 and the pitch period T 123 allows a pitch period starting point β
10 24 and a pitch period ending point 26 of the pitch period T 123
11 to be determined. The pitch period starting point β 24 and
12 ending point 26 define the boundaries of the pitch period T 123.
13 By using the centroid to determine the segmentation, the present
14 invention results in a balancing of the "weight" or left and
15 right moments of the pitch waveform samples around the centroid.

16 The segmentation of the waveform samples, in accordance with
17 the present invention, is preferably performed using a system 30
18 as illustrated in figure 3. An input analog speech signal 32,
19 such as a human voice that is to be compressed or modified, from
20 an input device 34, such as a microphone, is sampled by a
21 conventional analog-to-digital converter (ADC) 36 at a
22 conventional sampling rate suitable for speech processing.
23 Digitized speech samples 38 are provided to a conventional
24 computer system 40, such as a Sun workstation or a desktop

Serial No.:
Inventors: Kang and Fransen

PATENT APPLICATION
Navy Case No. 77,023

1 personal computer. The computer system 40 performs the
2 segmentation (as indicated in Fig. 4 - to be discussed) and any
3 analysis or processing required for speaker verification, speaker
4 recognition, text to speech, compression, modification,
5 synthesis, etc. The segmented waveform in modified or unmodified
6 form can be stored in a memory (disk or RAM - not shown) of the
7 system 40. If the waveform is being modified, such as when
8 disguised speech is to be produced, modified speech waveform 42
9 samples are converted by a conventional digital-to-analog
10 converter (DAC) 44 into an analog speech signal 46 and provided
11 to an output device 48, such as a speaker. The process of the
12 present invention can also be stored in a portable medium, such
13 as a disk, and carried from system to system.

14 The segmentation operation segments by determining the
15 centroid, which is performed by the computer system 40, as
16 illustrated in detail in figure 4. This segmentation operation
17 starts by generating a ramp R(.) 171 of one pitch period duration
18 in a ramp function generator 50. More specifically, the
19 generator 50 is responsive to a pitch period T 123 for generating
20 ramp R(.) 171, expressed by

$$\begin{array}{lcl} 21 & R(i) & = \begin{cases} i & \text{for } -T/2 \leq i \leq T/2 - 1 \\ 22 & \\ 23 & 0 & \text{elsewhere} \end{cases} \end{array} \quad (2)$$

24 The ramp R(.) 171 is then correlated in a correlator 52 with
25 low pass filtered speech samples S'(.) 126 to produce a centroid

Serial No.:
Inventors: Kang and Fransen

PATENT APPLICATION
Navy Case No. 77,023

function or the centroid histogram samples C(.) 182. The use of low pass filtered speech samples S'(.) 126 is preferred because it is free of high frequency information often present in the speech waveform. By definition, a centroid function is the sum of the products of the ramp R(.) 171 samples and the low-pass filtered speech samples S'(.) 126 with a successive mutual delay (which is a cross correlation function). Thus, the centroid histogram samples C(.) 182 are expressed by

$$\begin{aligned} C(i) &= \sum_{j=-T/2}^{T/2-1} R(j)S'(i+j) \quad L - T/2 \leq i \leq L + T/2 - 1 \\ &= \sum_{j=-T/2}^{T/2-1} jS'(i+j) \quad L - T/2 \leq i \leq L + T/2 - 1 \end{aligned} \quad (3)$$

where L is the midpoint of the centroid analysis frame. As noted from the above expression (3), the samples C(.) 182 are computed for one pitch period around the center of the analysis frame. The typical centroid histogram samples C(.) 182 waveform is illustrated in figure 2(b), which was previously discussed.

Next, a local minimum search 54 is performed on the samples C(.) 182 (see figure 4) to determine a local minimum location α . As previously noted, the midpoint of the pitch waveform coincides with the local minimum of the samples C(.) 182. This minimum location, denoted by α 199, is obtained from

$$\alpha = \min_i \{C(i)\} \quad L - T/2 \leq i \leq L + T/2 - 1 \quad (4)$$

As illustrated in figures 2(a) and 2(b), the minimum location α 199 corresponds to the midpoint 20 of the pitch period T 123. Thus, the pitch epoch begins at $\alpha - T/2$, the pitch period starting point β 24, and ends at $\alpha + T/2 - 1$ or pitch period ending point 26.

The minimum location α 199 needs refinement because the pitch period T 123 provided by a pitch tracker 121 (figure 9) is often not too accurate. Thus, both the local minimum location α 199 and pitch period T 123 are refined by repeating the above local minimum search 54 for each $T \pm \Delta T$ where ΔT is as much as $T/16$ (6.25% of T). This refinement 56 improves the segmentation performance. The refined local minimum location and refined pitch period are denoted by α' and T' , respectively.

Finally, segmented pitch waveform samples 131 are excised from the speech samples $S(.)$ 128 by a switch 58 from time $\alpha' - T'/2$ to time $\alpha' + T'/2 - 1$.

Once the segmented pitch waveform samples 131 are excised they can be modified, replicated, etc., as will be discussed in more detail later, to produce a reconstructed speech waveform. This is accomplished by replicating and concatenating pitch waveforms. Because the synthesis frame size M is generally greater than the pitch period T' , the segmented speech waveform

Serial No.:
Inventors: Kang and Fransen

PATENT APPLICATION
Navy Case No. 77,023

1 is usually replicated more than once. The segmented waveform is
2 always replicated in its entirety. Near the boundary of the
3 synthesis frame, it is necessary to decide whether the segmented
4 waveform of the current frame should be replicated again or the
5 segmented waveform of the next frame should be copied. The
6 choice is determined by the remaining space in relation to the
7 length of the segmented waveform T' . If the remaining space is
8 greater than $T'/2$, the segmented waveform of the current frame is
9 replicated again. On the other hand, if the remaining space is
10 less than or equal to $T'/2$, the segmented waveform of the next
11 frame is copied. Any significant discontinuity at either of the
12 segmented pitch waveform boundaries 24 and 26 (figure 2(a)) will
13 produce clicks or warbles in the reconstructed speech. To avoid
14 discontinuities, the system performs a three-point interpolation
15 at the pitch epoch (see FORTRAN program on pages A1 through A10
16 of the Appendix for details of this operation).

17 As noted previously the segmented pitch waveform samples 131
18 can be used for speaker verification, speaker recognition, text
19 to speech, compression, synthesis or modification of the speech
20 waveform. The modification operation can independently modify
21 the utterance rate, the pitch, and the resonance frequencies of
22 the original speech waveform.

23 Referring now to figures 5(a) - 5(c), the speech utterance
24 rate is altered by simply changing the number of pitch waveforms

Serial No.:
Inventors: Kang and Fransen

PATENT APPLICATION
Navy Case No. 77,023

1 replicated at the output. Therefore, the utterance rate is
2 controlled by synthesis frame size, M , relative to the analysis
3 frame size, N , which are internal parameters that can be altered
4 by the operator. The relationship of N and M are shown in figure
5 6. Three cases for the relationship between N and M are: (1) M
6 $= N$: In this case, the utterance rate is unaltered because the
7 same number of pitch waveforms are present in both the input and
8 output frames (see figure 5(a)). (2) $M > N$: The output speech is
9 slowed down by replicating the pitch waveform 60 more than once,
10 producing replicated waveforms 62 used to fill the synthesis
11 frame M (see figure 5(b) for an example). (3) $M < N$: The output
12 speech is sped up because the output frame M has fewer pitch
13 waveforms than the input frame N (see figure 5(c)). In these
14 examples of utterance rate change the pitch period and resonance
15 frequencies of the original speech are not affected by modifying
16 the speech rate.

17 Pitch can be changed by expanding or compressing the pitch
18 waveform. Alteration of the pitch period in this invention does
19 not affect the speech utterance rate, but resonant frequencies do
20 change in proportion to the pitch. It is common knowledge that
21 high-pitch female voices have higher resonant frequencies than
22 low-pitch male voices for the same vowel. The natural coupling
23 of pitch frequency and resonant frequencies is beneficial.

24 Figures 7(a)-7(c) illustrate the effect of changing the

Serial No.:
Inventors: Kang and Fransen

PATENT APPLICATION
Navy Case No. 77,023

1 pitch. Figure 7(a) shows the original speech. Figure 7(b) is
2 altered speech played back with a 30% lower pitch. Figure 7(c)
3 is altered speech played back with a 30% higher pitch.

4 The resonant frequencies of speech can be modified by
5 altering the pitch waveform spectrum. An example of such an
6 alteration is illustrated in figure 8. In step 1, a conventional
7 discrete Fourier transform (DFT) is applied to the segmented
8 pitch waveform samples 131 to produce an amplitude spectrum 74.
9 In step 2, the spectrum 74 is modified in some conventional
10 manner to produce a modified amplitude spectrum 78. For example,
11 the first resonant frequency in the spectrum 74 can be shifted to
12 the left as shown by the spectrum 78. In step 3, a conventional
13 Hilbert transformation is performed on spectrum 78 to produce a
14 modified phase spectrum 82. In step 4, an inverse discrete
15 Fourier transform (IDFT) is performed on amplitude spectrum 78
16 and phase spectrum 82 to produce a modified pitch waveform 135
17 with altered spectral characteristics. This waveform 135 can
18 then be used to generate speech.

19 Figure 9 shows the structural elements in the computer
20 system 40 of figure 3 for performing the operations of segmenting
21 and reconstructing a speech waveform. As shown in figure 9,
22 three inputs 111, 113 and 32 are provided: an analysis frame
23 size N 111 (an integer from 60 to 240), a synthesis frame size M
24 113 (an integer from 60 to 240) and the input analog speech

Serial No.:
Inventors: Kang and Fransen

PATENT APPLICATION
Navy Case No. 77,023

1 signal 32. The analysis frame size N 111 and the synthesis frame
2 size M 113 are provided by the operator prior to start up of
3 system 30 (figure 3). The analog signal 32 from an input device,
4 such as a microphone, is converted by the ADC 36 into a series of
5 digitized speech samples 38 supplied at an 8-kHz rate. Although
6 not shown, the analog signal 32 is low pass filtered by the ADC
7 36 prior to the conversion to pass only signals below 4 kHz. The
8 digitized speech samples 38 are conventionally filtered by a low-
9 pass filter 119 to pass low pass filtered speech samples 120 at
10 audio frequencies below about 1 kHz while the original signal is
11 delayed in a shift register 125 (to be discussed) to produce
12 delayed speech samples S(.) 128. The pitch of the low pass
13 filtered speech samples 120 is pitch period tracked by a
14 conventional pitch tracker 121 to produce the pitch period T 123
15 (figure 4). A conventional pitch tracker is described in Digital
16 Processing Of Speech Signals, by Rabiner et al, Prentice-Hall,
17 Inc., NJ 1978, Chapter 4. The low pass filtered speech samples
18 120 are delayed in a shift register 127 (to be discussed). The
19 delays of shift registers 125 and 127 are preselected to time
20 align the low pass filtered speech samples S'(.) 126 and speech
21 samples S(.) 128 with the pitch period T 123 for input to a
22 pitch-synchronous speech segmentor 129. The pitch period T 123,
23 the low pass filtered speech samples S'(.) 126, the speech
24 samples S(.) 128 and the analysis frame size N 111 are used to

Serial No.:
Inventors: Kang and Fransen

PATENT APPLICATION
Navy Case No. 77,023

1 perform segmentation in the segmentor 129 of the original signal
2 as described with respect to figure 4. The segmented pitch
3 waveform samples 131 are then transformed in an application
4 dependent pitch waveform transformator 133, in one or more of the
5 ways as previously discussed, to produce the modified pitch
6 waveform 135. Speech is reconstructed in a speech waveform
7 reconstructor 139 using the modified pitch waveform 135 and the
8 synthesis frame size M 113 to produce the modified speech
9 waveform 42. The modified speech waveform 42 is converted by DAC
10 44 into the output analog speech signal 46 which is supplied to
11 an output device, such as a speaker (not shown).

12 The operations of figure 9, including the segmentation of
13 figure 4, are described in more detail in the FORTRAN source code
14 Appendix included herein.

15 In a hardware embodiment of the invention, the pitch
16 synchronous segmentation of speech in the present invention can
17 also be performed by an exemplary system 148 using discrete
18 hardware components, as illustrated in figure 11. In the
19 exemplary system 148, the minimum location and pitch period
20 refinement 56 (figure 4) is not performed. Also, the analysis
21 frame size N 111 is restricted to the range where $160 \leq N \leq 240$.

22 Before figure 11 is discussed, reference will now be made to
23 figures 10(a) and 10(b). Figure 10(a) illustrates timing
24 circuits for generating the various timing signals used in the

Serial No.:
Inventors: Kang and Fransen

PATENT APPLICATION
Navy Case No. 77,023

1 system of figure 11, and figure 10(b) illustrates the control
2 timing signals generated by the timing circuits of figure 10(a).

3 In figure 10(a), a clock generator 136 generates eight mega
4 Hertz (8 MHz) clocks which are applied to an upper input of an
5 AND gate 138 and to a 1000:1 frequency count down circuit 140.
6 At this time the AND gate 138 is disabled by a 0 state signal
7 from the Q output of a flip flop 142. The 8 MHz clocks are
8 continuously counted down by the 1000:1 frequency count down
9 circuit 140 to generate an 8 kHz speech sampling clock A (shown
10 in figure 10(b)) each time that the count down circuit 140 counts
11 1000 8 MHz clocks and then is internally reset to zero (0) by the
12 1000 th 8 MHz clock. Note that the interpulse period of clock A
13 is 125 microseconds (μ s).

14 The 8 kHz speech sampling clock A is applied to M:1 and N:1
15 frequency count down circuits 144 and 146. It will be recalled
16 that the synthesis frame size M 113 and the analysis frame size N
17 111 are internal parameters that can be altered by the operator.
18 Thus, the values of M and N are selected by the operator.

19 The 8 kHz clock A is counted down by the M:1 frequency count
20 down circuit 144 to generate an 8 kHz/M synthesis frame clock C
21 (shown in figure 10 (b)) each time that the count down circuit
22 144 counts M 8 kHz A clocks and then is internally reset to 0 by
23 the M th 8 kHz clock. In a similar manner, the 8 kHz clock A is
24 counted down by the N:1 frequency count down circuit 146 to

Serial No.:
Inventors: Kang and Fransen

PATENT APPLICATION
Navy Case No. 77,023

1 generate an 8 kHz/N analysis frame clock B (shown in figure
2 10(b)) each time that the count down circuit 146 counts N 8 kHz A
3 clocks and then is internally reset to 0 by the Nth 8kHz clock.

4 The 8 kHz/N analysis frame clock B is also applied to a
5 25 μ s delay circuit 147 to produce a selected centroid histogram
6 samples transfer signal E (shown in figure 10 (b)) which occurs
7 25 μ s after each B clock. In a similar manner, the 8 kHz/N B
8 clock is applied to a 50 μ s delay circuit 150 to produce a begin
9 pitch waveform modification signal F (shown in figure 10 (b))
10 which occurs 50 μ s after the B clock. The B clock is also
11 applied to a 100 μ s delay circuit 152 to produce a ramp transfer
12 signal D (shown in figure 10 (b)) which occurs 100 μ s after the B
13 clock.

14 Each time that an F clock is generated by the 50 μ s delay
15 circuit 150, that F clock sets the flip flop 142 to cause the Q
16 output of the flip flop 142 to change to a 1 state output. This
17 1 state output enables the AND gate 138 to pass 8 MHz clocks.
18 These 8 MHz clocks from AND gate 138 will henceforth be called T₂
19 pulses G, which will be applied to a shift register 195 in figure
20 11 (to be discussed).

21 The T₂ pulses G from AND gate 138 are counted by a T₂:1
22 frequency count down circuit 154 to generate a T₂:1 signal each
23 time that the count down circuit 154 counts T₂ pulses and then is
24 internally reset to 0 by the T₂th 8 MHz clock that occurs after

Serial No.:
Inventors: Kang and Fransen

PATENT APPLICATION
Navy Case No. 77,023

1 the flip flop 142 is set. The $T_2:1$ clock also resets the flip
2 flop 142 so that the Q output of the flip flop 142 changes to a 0
3 state to disable the AND gate 138. Thus, no more T_2 pulses G are
4 supplied to the shift register 195 in figure 11 at this time.

5 As shown in figure 10(b), the T_2 8 MHz pulses G start with
6 the generation of the begin pitch waveform modification signal F
7 and terminate after the frequency count down circuit 154 has
8 counted T_2 8 MHz pulses G after the generation of the F signal.

9 Referring back to figure 11, the parameter analysis frame
10 size N 111 signal is applied to shift registers 183 and 191,
11 switches 187 and 193, minimum locator 189, and parallel-to-serial
12 shift register 195. Speech samples $S(.)$ 128 are fed at the time
13 of the A clocks through AND gate 155 to the shift register 191.
14 Pitch period T 123 signal is fed at the time of the D clocks
15 through AND gate 159 to a shift register 167 and a ramp generator
16 169. The low-pass filtered (LPF) speech samples $S'(.)$ 126 are
17 fed at the time of the A clock through AND gate 163 to a shift
18 register 165.

19 A conventional pitch tracker 121 (figure 9) used for this
20 embodiment is able to track pitch with a range of 51 Hz to 400
21 Hz. A low-pitch male voice of 51 Hz corresponds to a pitch
22 period, T, of 156 speech samples. A high-pitch female voice of
23 400 Hz corresponds to a pitch period, T, of 20 speech samples.
24 Thus, the segmentation process must be able to handle pitch

Serial No.:
Inventors: Kang and Fransen

PATENT APPLICATION
Navy Case No. 77,023

1 waveforms having 20 to 156 speech samples. Shift register 165
2 retains 156 filtered speech samples $S'(\cdot)$ 126, and shift register
3 175 stores ramp samples R_1 to R_T .

4 A ramp generator 169 develops an appropriate ramp $R(\cdot)$ 171
5 to be fed at the time of the ramp transfer signal D through an
6 AND gate 173 to the shift register 175. The number of ramp
7 samples transferred is T, and the appropriate ramp R_1 to R_T from
8 the following list is transferred.

9 R_1 to R_{20} : -10,-9,-8,...,-1,0,1,...,7,8,9
10 R_1 to R_{21} : -10,-9,-8,...,-1,0,1,...,8,9,10
11 R_1 to R_{22} : -11,-10,-9,...,-1,0,1,...,8,9,10
12
13
14
15 R_1 to R_{154} : -77,-76,-75,...,-1,0,1,...,74,75,76
16 R_1 to R_{155} : -77,-76,-75,...,-1,0,1,...,75,76,77
17 R_1 to R_{156} : -78,-77,-76,...,-1,0,1,...,75,76,77

18 Corresponding ramp samples R_1 to R_T from shift register 175
19 and corresponding low pass filter speech samples S'_1 to S'_T from
20 shift register 165 are respectively cross multiplied in
21 associated multipliers 166 to develop and apply cross products
22 179 to a summation unit 181.

23 Cross-products 179 of filtered speech samples S'_1 to S'_T
24 with ramp R_1 to R_T pass through the summation unit 181 to form
25 centroid histogram samples $C(\cdot)$ 182 for feeding into buffer 183.
26 Ramp R_1 to R_T remains fixed over an analysis frame of N speech
27 samples $S(\cdot)$ 128. An analysis frame of N filtered speech samples
28 $S'(\cdot)$ 126 produces a frame of N sums of cross products designated

Serial No.:
Inventors: Kang and Fransen

PATENT APPLICATION
Navy Case No. 77,023

1 C_1 to C_N in register 183. C_1 to C_N is also designated frame 1 in
2 register 183. Because a pitch waveform can spread over three
3 frames, selection of a pitch waveform progresses from the middle
4 of the three frames of register 183, at location $3N/2$. Location
5 $3N/2$ is positioned in the middle of frame 2 of register 183.
6 Because the search is now centered in frame 2, a one frame delay
7 is introduced in the segmentation process. Register 167 delays
8 the pitch one frame to properly line up the pitch in time with
9 frame 2 of register 183.

10 Analysis frame size N 111 samples is fixed prior to the
11 start up of ADC 36 and DAC 44 (ADC and DAC are shown in figures 3
12 and 9). Since N can range in the exemplary system 148 from 160
13 to 240 samples and the pitch period T 123 can range from 20 to
14 156 samples, three frames, $3N$, of centroid samples are preserved
15 in register 183.

16 The goal is to find a pitch waveform to associate with frame
17 2 of register 183. The beginning of the pitch cycle must be
18 found such that a replication and concatenation process to be
19 performed later will not create audible discontinuities. Each
20 sum of cross products C_1 to C_N from summation unit 181 that is
21 fed into register 183 is an indication of the center of gravity
22 of a pitch waveform. The midpoint of a new pitch waveform occurs
23 when the center of gravity is at a relative minimum.

24 A search window for locating the segmented pitch waveform

Serial No.:
Inventors: Kang and Fransen

PATENT APPLICATION
Navy Case No. 77,023

1 samples 131 (of Figs. 4 and 11) is centered about the middle of
2 frame 2. A search controller 197, such as a microprocessor,
3 computes $\Delta = T_2/2$. The range of the search window is from centroid
4 histogram sample $C_{3N/2-\Delta}$ to $C_{3N/2+\Delta}$ which encompasses $2\Delta+1$ samples,
5 or a little greater than a pitch period of samples.

6 Once per analysis frame, centroid samples $C_{3N/2-\Delta}, \dots, C_{3N/2+\Delta}$
7 are fed at the time of the E signal through AND gates 185 and
8 through switch 187 to the minimum locator 189. Locator 189 is a
9 conventional device, such as a microprocessor, used for finding
10 the location of the minimum value of the centroid samples $C_{3N/2-\Delta}, \dots, C_{3N/2+\Delta}$
11 within the locator 189. The pitch period starting
12 point β 24 of the selected pitch waveform is in the range of
13 $3N/2-\Delta$ to $3N/2+\Delta$. The starting point β 24 is passed to the
14 switch 193. Switch 193 transfers T_2 speech samples from shift
15 register 191 to shift register 195. Segmented pitch waveform
16 samples 131 are available for the application dependent pitch
17 waveform transformator 133. Shift register 191 has a size of $6N$
18 to have sufficient speech samples available.

19 Figure 12 shows a first type of circuit for utilizing the
20 segmented pitch waveform samples 131 output of figure 11 to
21 modify the waveform spectrum. In this circuit of figure 12,
22 resonant frequencies of the segmented pitch waveform 131 are
23 altered. The application of timing signal G (figures 10(a) and
24 10(b)) to the shift register 195 (figure 11) enables segmented

Serial No.:
Inventors: Kang and Fransen

PATENT APPLICATION
Navy Case No. 77,023

1 pitch waveform samples 131 to be fed from the shift register 195
2 to a DFT unit 205. Amplitude and phase spectrum output 207 from
3 DFT unit 205 are changed by an amplitude and phase spectrum
4 modification unit 209 in a manner similar to that previously
5 described in figure 8.

6 To explain this amplitude and phase spectrum modification
7 being performed by circuit 209 of figure 12 reference will now be
8 made back to the description of figure 8.

9 The resonant frequencies of speech can be modified by
10 altering the pitch waveform spectrum. An example of altering the
11 first resonant frequency is illustrated in figure 8. In step 1,
12 a conventional DFT is applied to the segmented pitch waveform
13 samples 131 to produce the amplitude spectrum 74. In step 2, the
14 spectrum 74 is modified in some conventional manner to produce
15 the modified amplitude spectrum 78. For example, the first
16 resonant frequency in the spectrum 74 can be shifted to the left
17 as shown by the spectrum 78. In step 3, a conventional Hilbert
18 transformation is performed on spectrum 78 to produce the
19 modified phase spectrum 82. In step 4, an IDFT is performed on
20 amplitude spectrum 78 and phase spectrum 82 to produce the
21 modified pitch waveform 135 with altered spectral
22 characteristics. This waveform 135 can then be used to generate
23 speech. This would tend to disguise speaker identity.

24 Now referring back to figure 12, a modified amplitude

Serial No.:
Inventors: Kang and Fransen

PATENT APPLICATION
Navy Case No. 77,023

1 spectrum and phase spectrum signal 210 from the amplitude and
2 phase spectrum modification unit 209 is inverted using an IDFT
3 unit 211 and the resultant modified pitch waveform 135 is output
4 to a 156 sample serial-to-parallel shift register 213.

5 Figure 12 can be changed to pass the segmented pitch
6 waveform samples 131 unaltered through the circuit of figure 12
7 by removing the amplitude and phase spectrum modification circuit
8 209 and applying the output of the DFT unit 205 directly to the
9 input of the IDFT unit 211 or by applying the output from shift
10 register 195 (figure 11) directly to the input of shift register
11 213.

12 Another alternate embodiment of the discrete component
13 version of this invention is illustrated in figure 13. The
14 segmented pitch waveform samples 131 stored in shift register 195
15 pass through a stretching or shrinking transformation. Pitch
16 waveform samples 131 are applied to a DAC 321 with the 8 kHz
17 clock A (clock A generation is shown in figures 10(a) and 10(b)).
18 The analog pitch waveform is resampled by an ADC 323 at a new
19 sampling rate denoted by H (permissible values for H are $4 \text{ kHz} \leq$
20 $H \leq 16 \text{ kHz}$) to create the modified pitch waveform 135 with T"
21 samples stored in the shift register 213. Shrinking the pitch
22 waveform raises the pitch, and expanding the pitch waveform
23 lowers the pitch.

24 A discrete component waveform reconstruction circuit is

Serial No.:
Inventors: Kang and Fransen

PATENT APPLICATION
Navy Case No. 77,023

1 illustrated in figure 14. This circuit comprises the shift
2 register 213, a 156-sample, serial-to-parallel shift register
3 433, and two 156-sample, parallel-to-serial shift registers 431
4 and 435. Since the pitch period T_{123} has a range of 20 to 156
5 samples, each of the 156-sample registers 213, 431, 433, and 435
6 enable the storage of the maximum number of samples in a pitch
7 waveform.

8 A control circuit 445 generates $312 - T_2$ pulses at an 8 MHz
9 rate beginning at the time that clock E is generated. The
10 control circuit 445 includes a flip flop 441 which is enabled by
11 clock E to allow 8 MHz pulses to pass through an AND gate 437. A
12 frequency count down circuit 439 permits $312 - T_2$ 8 MHz pulses to
13 pass through the AND gate 437 before it counts to a count of
14 $312 - T_2$. When the frequency count down circuit 439 reaches a
15 count of $312 - T_2$, it resets the flip flop 441 and internally
16 resets itself to a 0 count. When reset, the Q output of the flip
17 flop 441 changes to a 0 state to disable the AND gate 437. At
18 this time no further 8 MHz pulses can be output from the control
19 circuit 445 until the flip flop 441 is reset by the next enabling
20 E clock.

21 Modified pitch waveform 135 samples are updated once per
22 analysis frame. For purposes of this description, the updating
23 operation of figure 14 will be described in relation to the
24 utilization circuit of figure 12. However, it should be

Serial No.:
Inventors: Kang and Fransen

PATENT APPLICATION
Navy Case No. 77,023

1 understood that a similar description of figure 14 is also
2 applicable to the utilization circuit of figure 13.

3 In operation, modified pitch waveform 135 samples from
4 figure 12 are serially clocked into serial-to-parallel register
5 213 by the G clock (figure 10(b)), which G clock is comprised of
6 T_2 8 MHz clocks. At the time of the B clock, the stored samples
7 in register 213 are shifted into and stored in parallel in the
8 parallel-to-serial shift register 431. Since T_2 is often less
9 than the 156-sample register-capacity of each of the registers
10 213 and 431, there is null data (i.e., data not related to the
11 pitch waveform) comprising $156 - T_2$ samples positioned in time
12 prior to the pitch waveform in the registers 213 and 431.

13 At the time of the next E clock, following the G clock
14 during which the modified pitch waveform 135 samples were stored
15 in the register 213, the flip flop 441 is set to enable AND gate
16 437 to pass 8 MHz clocks to registers 431 and 433. These 8 MHz
17 clocks from AND gate 437 enable the samples stored in the
18 register 431 to be serially clocked out of the register 431 into
19 register 433. This transfer repositions the null data in time
20 behind the speech data in register 433. More specifically, the
21 first 156 clock pulses from the AND gate 437 in the circuit 445
22 transfer the entire contents of the register 431 to register 433,
23 and the additional $156 - T_2$ clock pulses eliminate null data prior
24 to the speech data in register 433.

Serial No.:
Inventors: Kang and Fransen

PATENT APPLICATION
Navy Case No. 77,023

1 The 8 MHz clocks from the AND gate 437 are also counted by a
2 frequency count down circuit 439. When the circuit 439 reaches a
3 count of $(312 - T_2)$ 8 MHz clocks, it generates a signal to reset
4 the flip flop 441 to disable the AND gate 437 so that no further
5 8 MHz clock pulses are output from the control circuit 445 until
6 the flip flop 441 is set by the next enabling clock E.

7 The 8 kHz clock A is fed to a frequency count down circuit
8 443 to transfer in parallel the contents of register 433 to
9 register 435 and to internally reset the counter 443 to zero (0)
10 when the counter 443 has counted T_2 A clocks. Finally, T_2
11 samples of register 435 are fed at an 8 kHz rate by clock A to
12 form the waveform 42 which is then applied to the DAC 44 at the A
13 clock rate. The entire pitch waveform comprised of T_2 samples
14 must be transferred in its entirety. The resulting analog speech
15 signal 46 is then applied to the output device 48.

16 Additional details of uses for the present invention can be
17 found in Naval Research Laboratory report NRL/FR/5550-94-9743
18 entitled Speech Analysis and Synthesis Based on Pitch-Synchronous
19 Segmentation of the Speech Waveform by the inventors Kang and
20 Fransen, published November 9, 1994 and available from Naval
21 Research Laboratory, Washington, D.C. 20375-5320 and incorporated
22 by reference herein.

23 The present invention is described with respect to
24 performing pitch synchronous segmentation using centroid

Serial No.:
Inventors: Kang and Fransen

PATENT APPLICATION
Navy Case No. 77,023

1 analysis, however, the segmentation can be performed in other
2 ways. A direct approach is a method that determines pitch epochs
3 directly from the waveform. An example of such an approach is
4 peak picking in which the peaks 500 of the pitch waveforms are
5 used to find the segment speech waveform. For certain speech
6 waveforms, such an approach is feasible because the speech
7 waveform shows pitch epochs rather clearly as in figure 15. One
8 should be warned, however, that many speech waveforms do not show
9 pitch epochs clearly. This is particularly true with non-
10 resonant high-pitch female voices. As a result, this approach is
11 not preferred.

12 Contrary to the direct method which uses instantaneous
13 values of speech samples, a correlation method makes pitch epoch
14 determination based on the ensemble averaging of a certain
15 function derived from the speech waveform. The centroid method
16 presented previously is a correlation process. The concept of
17 the centroid originated in mechanical engineering to determine
18 the center of gravity of a flat object. The concept of the
19 center of gravity has been used in the field of signal analysis
20 in recent years (See Papoulis A, Signal Analysis, McGraw-Hill
21 Book Company, New York, New York 10017). For the speech
22 waveform, the quantity x is a time variable, $f(x)$ is the speech
23 waveform, x_1 is the pitch epoch, and $x_2 - x_1$ is the current pitch
24 period which is known beforehand. As elaborated in NRL Report

Serial No.:
Inventors: Kang and Fransen

PATENT APPLICATION
Navy Case No. 77,023

9743 (previously referenced), the above expression produces
virtually identical pitch epoch locations as the following
simplified expression:

$$\eta = \int_{x_1}^{x_2} x f(x) dx \quad \text{for } x_1 \leq x \leq x_2. \quad (6)$$

Thus, the centroid function is a cross correlation function
between a ramp function and $f(x)$. Ramp $R(\cdot)$ 171, as illustrated
in figure 16, appearing in the above equation is odd-symmetric
with respect to its midpoint. Other odd symmetric functions,
such as a sine function 512 and a step function 514 of figure 16,
can be used as a substitute for the ramp function. However,
these alternative functions do not work as well as the ramp
function and are thus not preferred.

The advantages of the present invention include the
following. Speech utterance rate can be changed without altering
the pitch or resonant frequencies. Pitch can be changed without
altering the utterance rate. Resonant frequencies can be changed
by spectrally shaping the pitch waveform without altering the
utterance rate or pitch. The modified speech is similar to the
original speech (not synthetic speech). Thus, the transformed
speech intelligibility and quality are excellent. This invention
has the feature of segmenting the speech waveform in terms of the

Serial No.:
Inventors: Kang and Fransen

PATENT APPLICATION
Navy Case No. 77,023

1 pitch waveform. In the invention, the pitch waveform is a
2 minimum inseparable entity of the speech waveform. Modification
3 of the pitch waveform leads to speech characteristic alteration.

4 The many features and advantages of the invention are
5 apparent from the detailed specification and, thus, it is
6 intended to cover all such features and
7 advantages of the invention which fall within the true spirit and
8 scope of the invention. Further, since numerous modifications
9 and changes will readily occur to those skilled in the art, it is
10 not desired to limit the invention to the exact construction and
11 operation illustrated and described, and accordingly all suitable
12 modifications and equivalents may be resorted to, falling within
13 the scope of the invention.

Serial No.:
Inventors: Kang and Fransen

PATENT APPLICATION
Navy Case No. 77,023

ABSTRACT OF THE DISCLOSURE

A system that synchronously segments a speech waveform using pitch period and a center of the pitch waveform. The pitch waveform center is determined by finding a local minimum of a centroid histogram waveform of the low-pass filtered speech waveform for one pitch period. The speech waveform can then be represented by one or more of such pitch waveforms or segments during speech compression, reconstruction or synthesis. The pitch waveform can be modified by frequency enhancement/filtering, waveform stretching/shrinking in speech synthesis or speech disguise. The utterance rate can also be controlled to speed up or slow down the speech.

Appendix
(FORTRAN Source Code)

```
c      NOTE   This program segments the speech waveform pitch
c            synchronously.  The segmented pitch waveform is
c            replicated and concatenated to generate
c            continuous speech.  The analysis frame size N
c            and synthesis frame size M are user specified.
c            Speech can be sped up by making N>M, or speech
c            may be slowed down by making N<M.
c
c      integer T,Tprime,ubnd
c      integer*2 is(240),idc(240),ilpf(240),ix(1)
c      integer*2 i5dc(1200),i5lpf(1200)
c      dimension amp(80),amp1(80),amp2(80),ampi(80)
c      dimension phase(80),phase1(80),phase2(80),phasei(80)
c      dimension pps(160),pw(160),xx(160)
c      character*75 fname
c
c      ----- voice i/o setup -----
c
c      write(6,1000)
c      1000  format('enter input speech file/')
c      read(5,1001) fname
c      1001  format(a)
c
c      *** initialize input device (not shown)
c
c      write(6,1002)
c      1002  format('enter output speech file/')
c      read(5,1001) fname
c
c      *** initialize output device (not shown)
c
c      ----- initialization -----
c
c      *** analysis frame size N:   60<=N<=240
c
c      N=100
c
c      *** synthesis frame size M:  60<=M<=240
c
c      M=100
```

```

C
C      *** constants
C
      lpfoffset=9
      twopi=2.*3.14159
C
C      ----- input speech samples -----
C
C      *** transfer N speech samples into array is(.)
C      *** in indicates how many samples actually transferred
C      *** subroutine spchin is not shown
C
100    call spchin(N,is,in)
      if(in.eq.0) go to 999
      ifrmct=ifrmct+1
C
C      ===== preprocessing =====
C
C      ----- remove dc from speech -----
C
      do 110 i=1,N
      x=is(i)
      call dcremove(x,y)
110    idc(i)=y
C
C      ----- store 5 dc-removed frames -----
C
      do 120 i=1,N
      i5dc(i)=i5dc(i+N)
      i5dc(i+N)=i5dc(i+2*N)
      i5dc(i+2*N)=i5dc(i+3*N)
      i5dc(i+3*N)=i5dc(i+4*N)
120    i5dc(i+4*N)=idc(i)
C
C      ----- low-pass filter -----
C
      do 130 i=1,N
      x=idc(i)
      call lpf(x,y)
130    ilpf(i)=y

```

```

C      ----- store 5 low-passed frames -----
C
C      do 140 i=1,N
C      i5lpf(i)=i5lpf(i+N)
C      i5lpf(i+N)=i5lpf(i+2*N)
C      i5lpf(i+2*N)=i5lpf(i+3*N)
C      i5lpf(i+3*N)=i5lpf(i+4*N)
140    i5lpf(i+4*N)=ilpf(i)
C
C      ===== analysis =====
C
C      ----- pitch tracker -----
C
C      NOTE    Use any reliable pitch tracker with an internal
C              two frame delay (pitch tracker not shown)
C
C      call pitch(N,i5lpf,T)
C      if(T.gt.128) T=128
C
C      ----- upper and lower bounds of search window -----
C
C      icenter=2.5*N
C      if(icenter.lt.T) icenter=T
C      lbnd=icenter-.5*(T+1)
C      ubnd=icenter+.5*(T+1)
C
C      ----- find pitch epoch and refine -----
C
C      call centroid(lbnd,ubnd,T,i5lpf,small,loc)
C      call adjust(T,loc,i5lpf,small,sadj,locadj,Tprime)
C
C      ----- compensate for lpf delay -----
C
C      locadj=locadj-lpfoffset
C
C      ----- extract one pitch-waveform and compute rms -----
C
C      index=locadj-Tprime/2
C      if(index.ge.1) go to 150
C      index=1
150    k=0
C      sum=0.
C      do 160 i=index,index+Tprime-1
C      k=k+1
C      pps(k)=i5dc(i)
160    sum=sum+pps(k)**2
C      rms=sqrt(sum/Tprime)

```

```

C
C      NOTE    Introduce pitch modification here (expand or
C              compress pps(.) and change Tprime accordingly)
C
C      ----- Fourier transform the extracted pitch waveform-----
C
C      NOTE    The pitch waveform is interpolated in the
C              frequency domain during the intra-pitch period
C
C      call dft(Tprime,pps,amp,phase,nn)
C
C      NOTE    Introduce spectrum modification here
C
C      do 170 i=nn+1,80
C      amp(i)=0.
170  phase(i)=0.
C
C      ----- store two frames of data -----
C
C      *** amplitude spectrum of pitch waveform
C
C      do 180 i=1,80
C      amp2(i)=amp1(i)
180  amp1(i)=amp(i)
C
C      *** phase spectrum of pitch waveform
C
C      do 181 i=1,80
C      phase2(i)=phase1(i)
181  phase1(i)=phase(i)
C
C      *** pitch period
C
C      ipt2=ipt1
C      ipt1=Tprime
C
C      *** pitch waveform rms
C
C      irms2=irms1
C      irms1=rms

```

```

C
C      ----- interpolation rate -----
C
C      NOTE    Use a faster interpolation if rms changes
C              significantly across frame boundary
C
C      ratio=iabs(irms1-irms2)
C      if(ratio.le.3.) ur=1.
C      if(ratio.gt.3.and.ratio.le.6) ur=1.2
C      if(ratio.gt.6) ur=1.4
C
C      ===== synthesizer =====
C
C      do 300 l=1,M
C
C      if(im-ipti)240,200,200
200  im=0
C      ----- pitch epoch -----
C
C      NOTE    At each pitch epoch, amplitude normalize
C              the pitch waveform of the previous pitch
C              period and dump out sample by sample.
C
C      *** amplitude normalization factor
C
C      sum=0.
C      do 210 i=1,ipti
210  sum=sum+xx(i)**2
C      gain=rmsi/sqrt(sum/ipti)
C
C      *** amplitude normalize past pitch waveform
C
C      do 220 i=1,ipti
C      u3=u2
C      u2=u1
C      u1=gain*xx(i)
C
C      *** perform 3-point interpolation only at pitch epoch
C
C      u0=u2
C      if(i.eq.2) u0=.25*u3+.5*u2+.25*u1
C
C      *** dump out sample by sample
C
C      if(u0.gt.32767.) u0=32767.
C      if(u0.lt.-32767.) u0=-32767.
C      ix(1)=u0
C

```

```

c      *** output one speech sample from array ix(.)
c      *** subroutine spchout is not shown
c
220    call spchout(1,ix)
c
c      *** interpolation factor
c
      factor=ur*1/float(M)
      if(factor.gt.1.) factor=1.
c
c      *** rms interpolation
c
      rmsi=irms2+factor*(irms1-irms2)
c
c      *** pitch interpolation
c
      ipti=ipt2+factor*(ipt1-ipt2)
c
c      *** amplitude spectrum interpolation
c
      do 230 i=1,80
230    ampi(i)=amp2(i)+factor*(amp1(i)-amp2(i))
c
c      *** phase spectrum selection
c
      if(factor.gt..5) go to 235
      do 232 i=1,80
232    phasei(i)=phase2(i)
      go to 238
c
235    do 236 i=1,80
236    phasei(i)=phase1(i)
c
c      ----- inverse discrete Fourier transform -----
c
238    call idft(ipti,ampi,phasei,pw)
c
c      ----- if not pitch epoch -----
c
240    im=im+1
      xx(im)=pw(im)
300    continue
      go to 100
c
c      -----
c
999    end

```

```

C
C      ===== subroutines =====
C
C      ----- dc remove subroutine -----
C
C      subroutine dcremove(a,b)
C
C      b=(a-a1)+.9375*b1
C      a1=a
C      b1=b
C      if(b.gt.32767.) b=32767.
C      if(b.lt.-32767.) b=-32767.
C      return
C      end
C
C      ----- low-pass filter subroutine (-3 db at 1025 hz) -----
C
C      subroutine lpf(r1,r2)
C
C      y19=y18
C      y18=y17
C      y17=y16
C      y16=y15
C      y15=y14
C      y14=y13
C      y13=y12
C      y12=y11
C      y11=y10
C      y10=y9
C      y9=y8
C      y8=y7
C      y7=y6
C      y6=y5
C      y5=y4
C      y4=y3
C      y3=y2
C      y2=y1
C      y1=r1
C      r2=.010*(y1+y19)+.013*(y2+y18)+.001*(y3+y17)-.024*(y4+y16)
&      -.045*(y5+y15)-.030*(y6+y14)+.039*(y7+y13)+.147*(y8+y12)
&      +.247*(y9+y11)+.285*y10
C      if(r2.gt.32767.) r2=32767.
C      if(r2.lt.-32767.) r2=-32767.
C      return
C      end

```

```

C
C      ----- pitch epoch finding subroutine -----
C
      subroutine centroid(i1,i2,ipp,i5lpf,small,loc)
      integer*2 i5lpf(1200)
C
      small=1000000.
      do 110 i=i1,i2
      sum=0.
      do 100 j=-ipp/2,-ipp/2+ipp-1
100    sum=sum+j*i5lpf(i+j)
      if(sum.gt.small) go to 110
      small=sum
      loc=i
110    continue
      return
      end
C
C      ----- pitch epoch refinement subroutine -----
C
      subroutine adjust(ipp,loc,i5lpf,small,sadj,locadj,ippadj)
      integer*2 i5lpf(1200)
C
      locadj=0
      Tprime=0
      sadj=1000000.
      irng=ipp/16
      do 110 i=loc-irng,loc+irng
      do 110 k=-irng,irng
      sum=0.
100    do 100 j=-(ipp+k)/2,-(ipp+k)/2+(ipp+k)-1
      sum=sum+j*i5lpf(i+j)
      if(sum.gt.sadj) go to 110
      sadj=sum
      locadj=i
      ippadj=ipp+k
110    continue
      return
      end

```

```

c
c      ----- discrete Fourier transform -----
c
      subroutine dft(ns,e1,amp,phase,nn)
      dimension e1(160),amp(80),phase(80)
c
      if(mod(ns,2).eq.0) nn=ns/2+1
      if(mod(ns,2).eq.1) nn=(ns+1)/2
      p=2.*3.1415926/ns
      tpi=2.*3.1415926
      tpit=tpi*(1./8000.)
      fs=8000./ns
c
100    do 110 j=1,nn
        rsum=0.
        xsum=0.
        const=tpit*fs*(j-1)
        do 120 i=1,ns
            arg=const*(i-1)
            rsum=rsum+e1(i)*cos(arg)
            xsum=xsum+e1(i)*sin(arg)
120        continue
        r=rsum/ns
        x=xsum/ns
        amp(j)=sqrt(r**2+x**2)
        phase(j)=atan2(x,r)
110    continue
      return
      end

```

```

C      ----- inverse discrete Fourier transform -----
C
C      subroutine idft(ns,amp,phase,e2)
C      dimension e2(160),amp(80),phase(80)
C
C      if(mod(ns,2).eq.0) nn=ns/2+1
C      if(mod(ns,2).eq.1) nn=(ns+1)/2
C      p=2.*3.1415926/ns
C      tpi=2.*3.1415926
C      tpit=tpi*(1./8000.)
C      fs=8000./ns
C
C      amp(1)=.5*amp(1)
C      if(mod(ns,2).eq.0) amp(nn)=.5*amp(nn)
C      do 210 i=1,ns
C      tsum=0.
C      const=tpit*fs*(i-1)
C      do 220 j=1,nn
C      arg=const*(j-1)
C      tsum=tsum+amp(j)*cos(arg-phase(j))
220    continue
C      e2(i)=2*tsum
210    continue
300    return
C      end

```

Navy Case No. 77,023

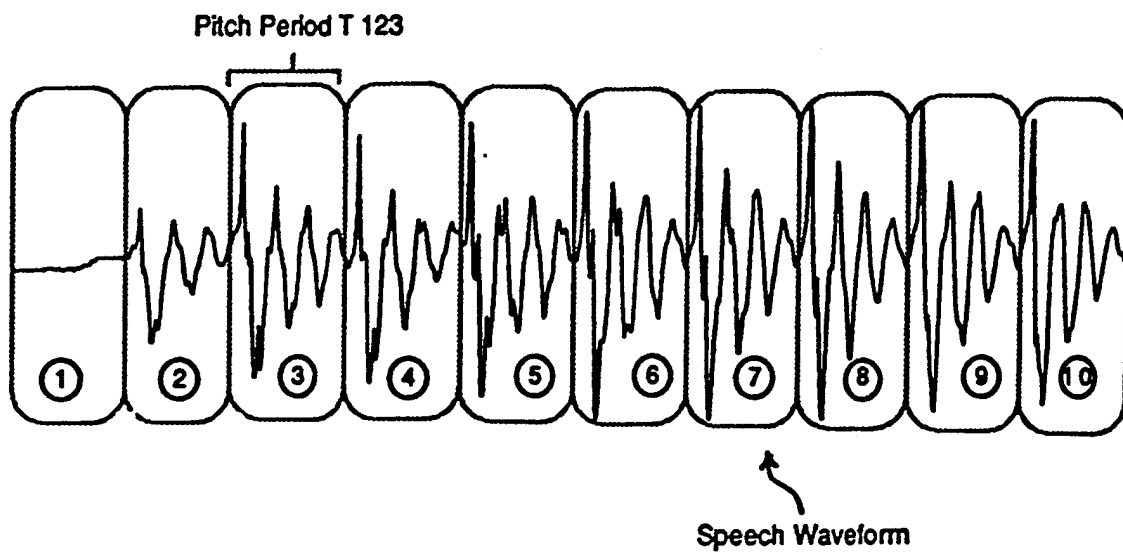


FIG. 1

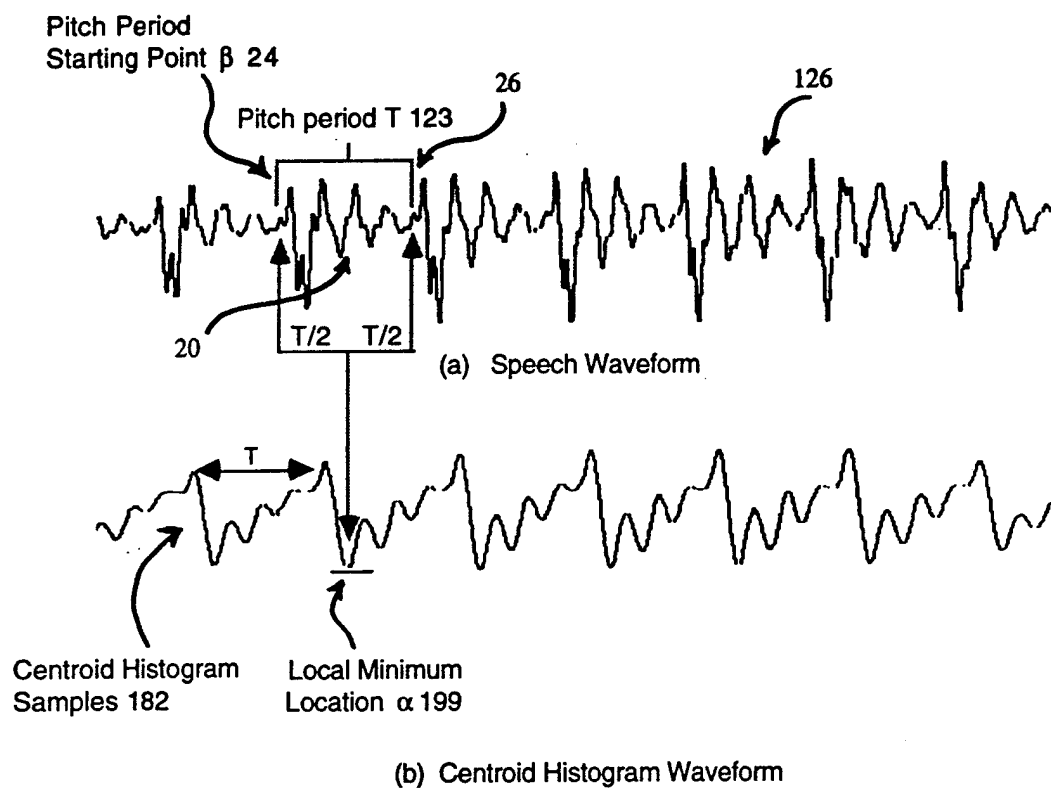


FIG. 2

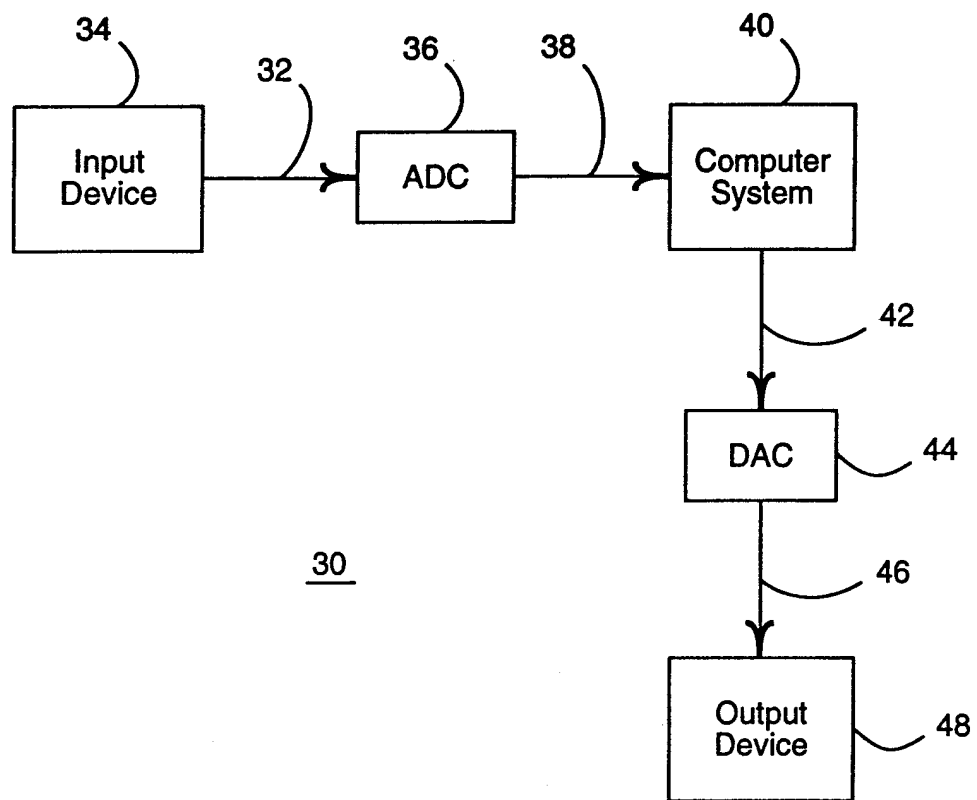


FIG. 3

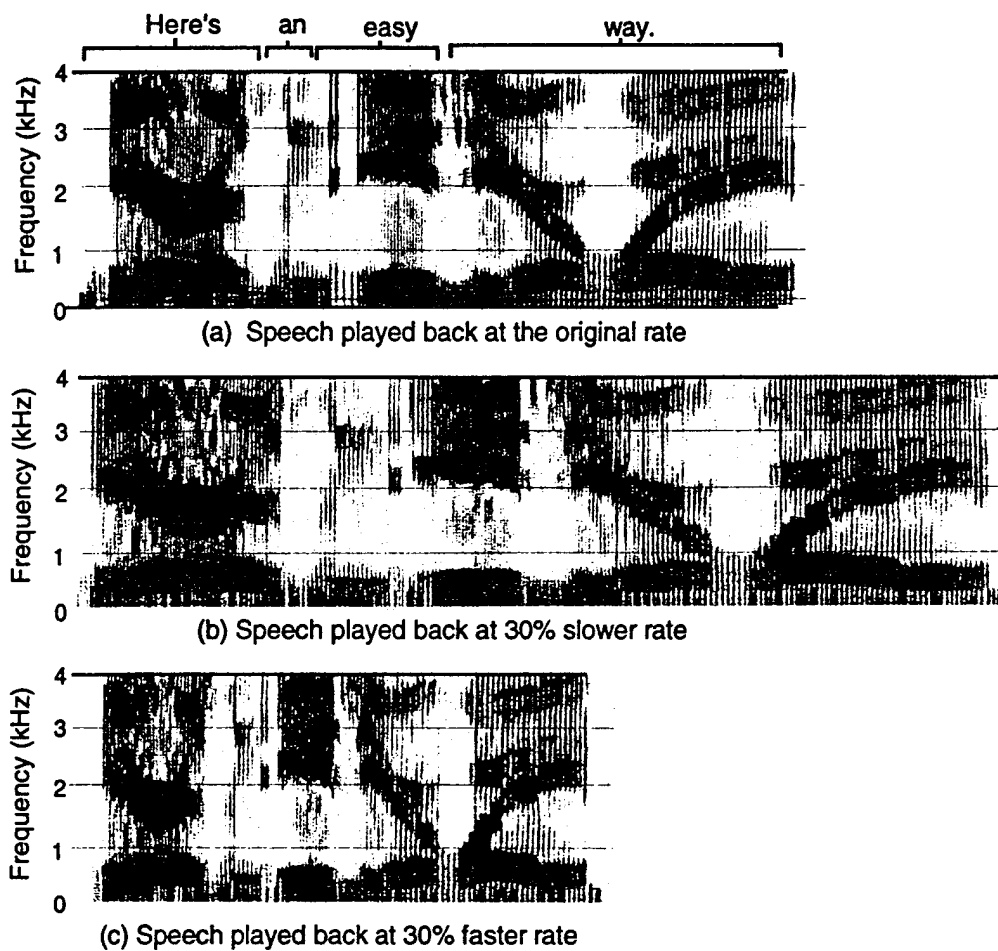


FIG. 5

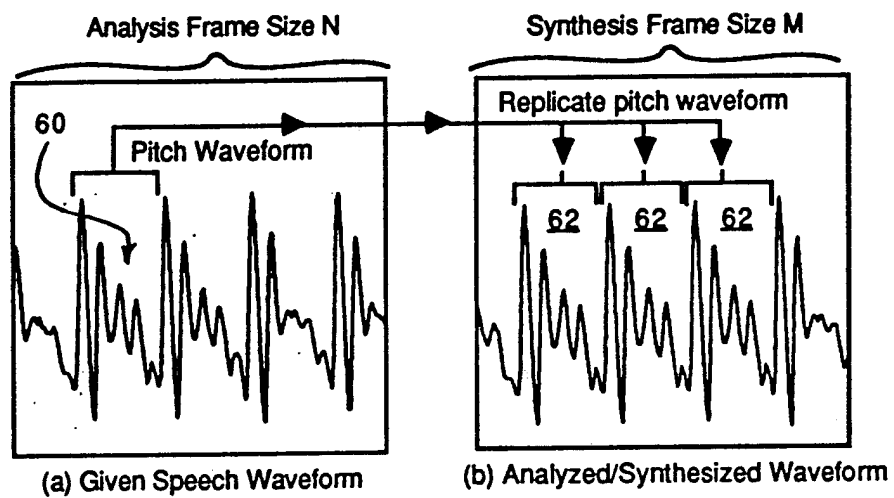


FIG. 6

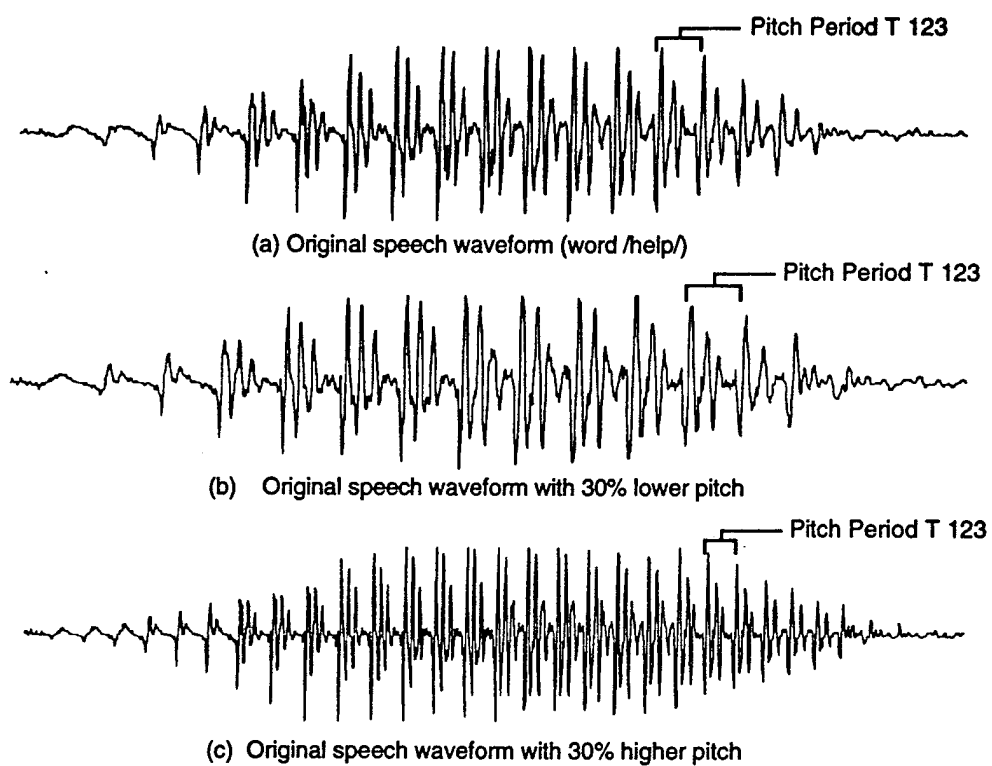


FIG. 7

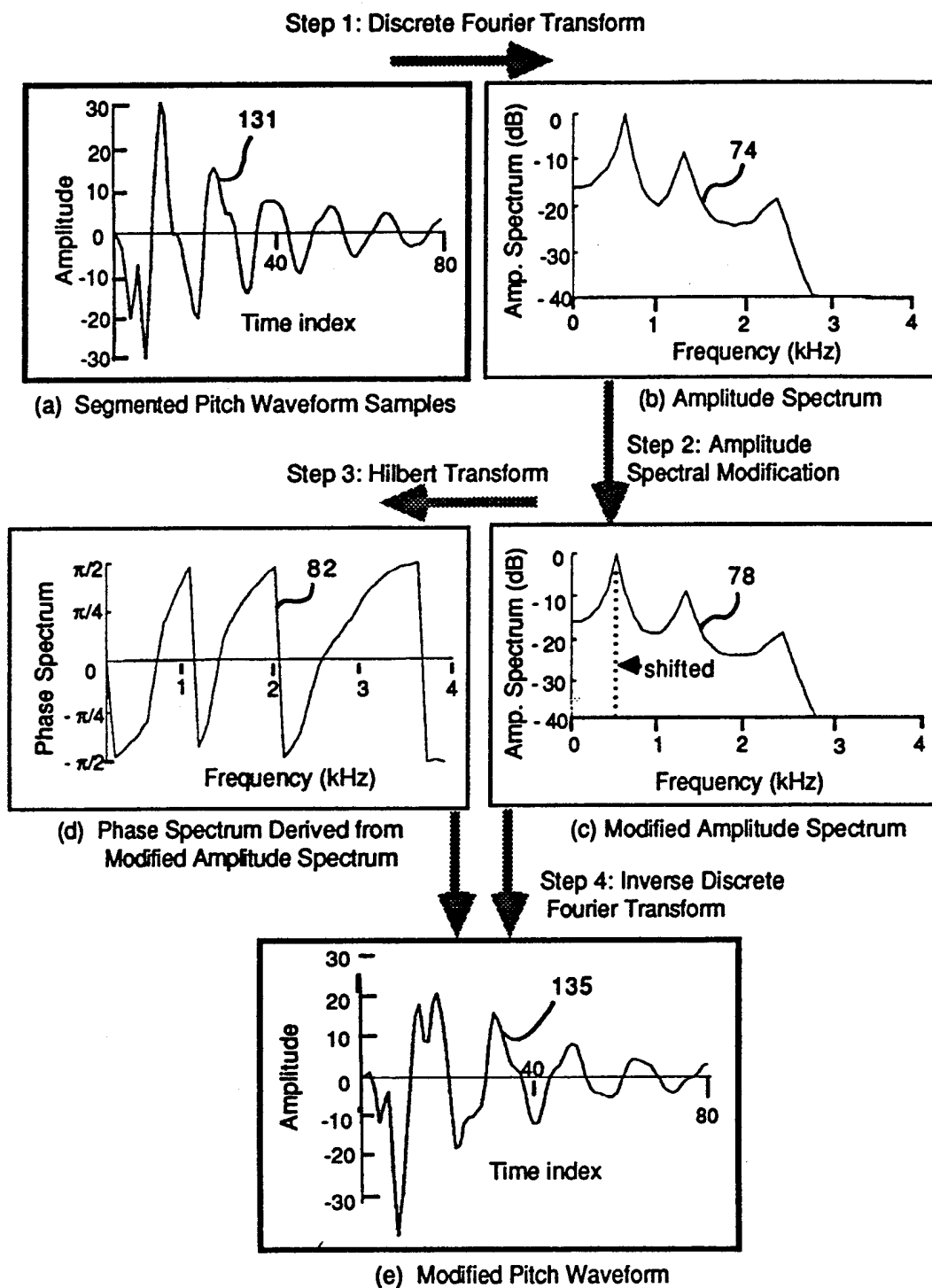


FIG. 8

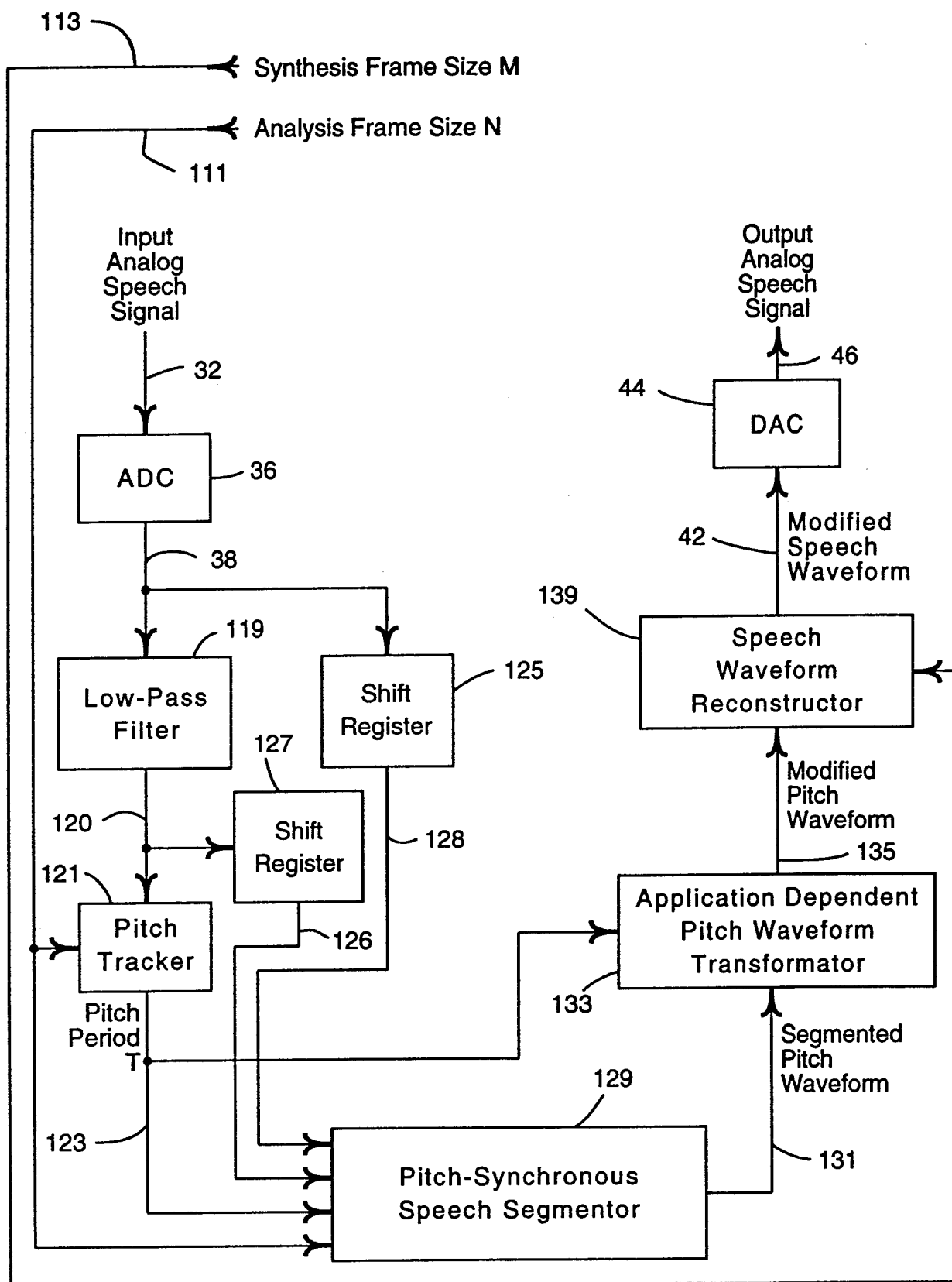


FIG. 9

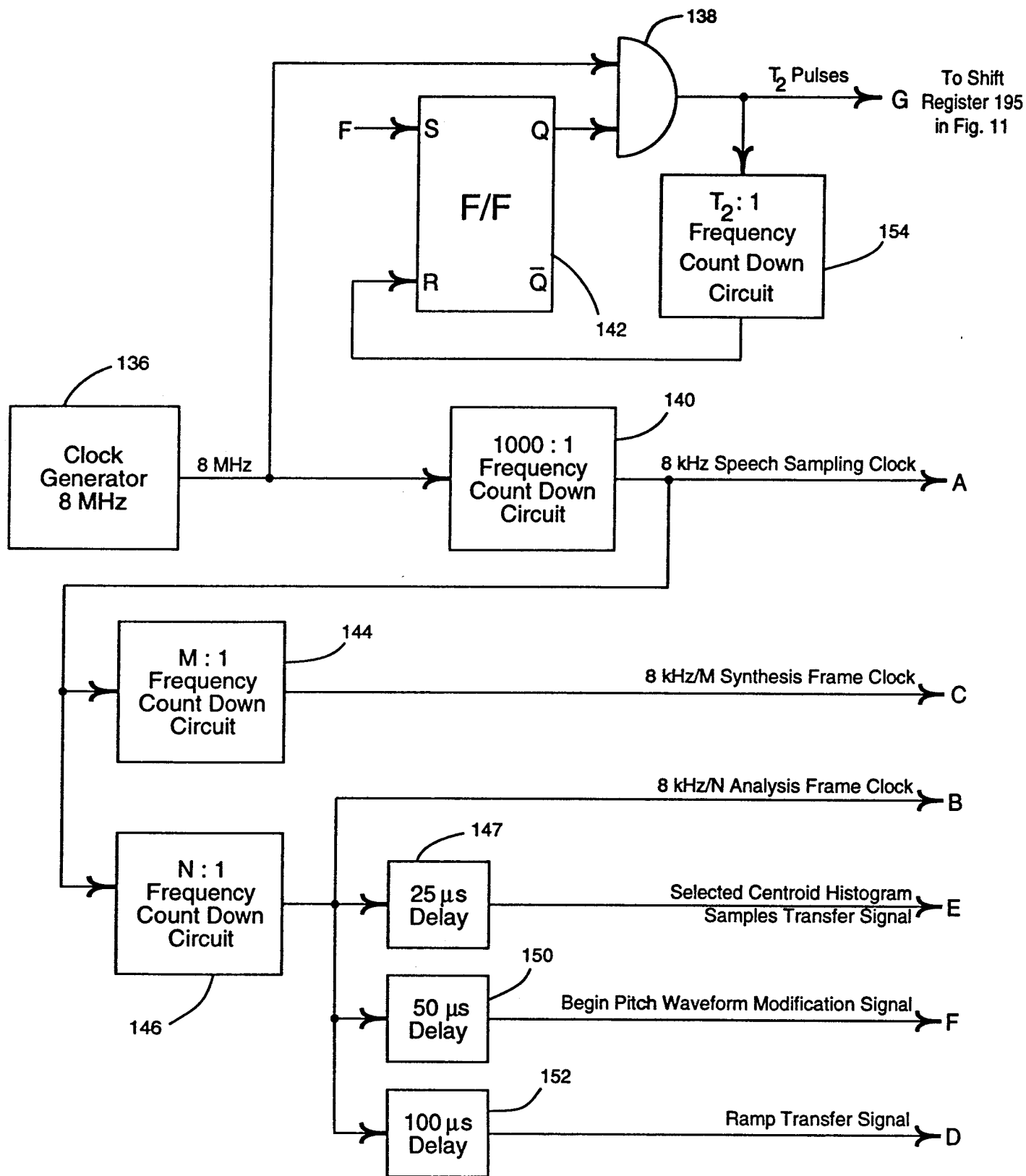


FIG. 10(a)

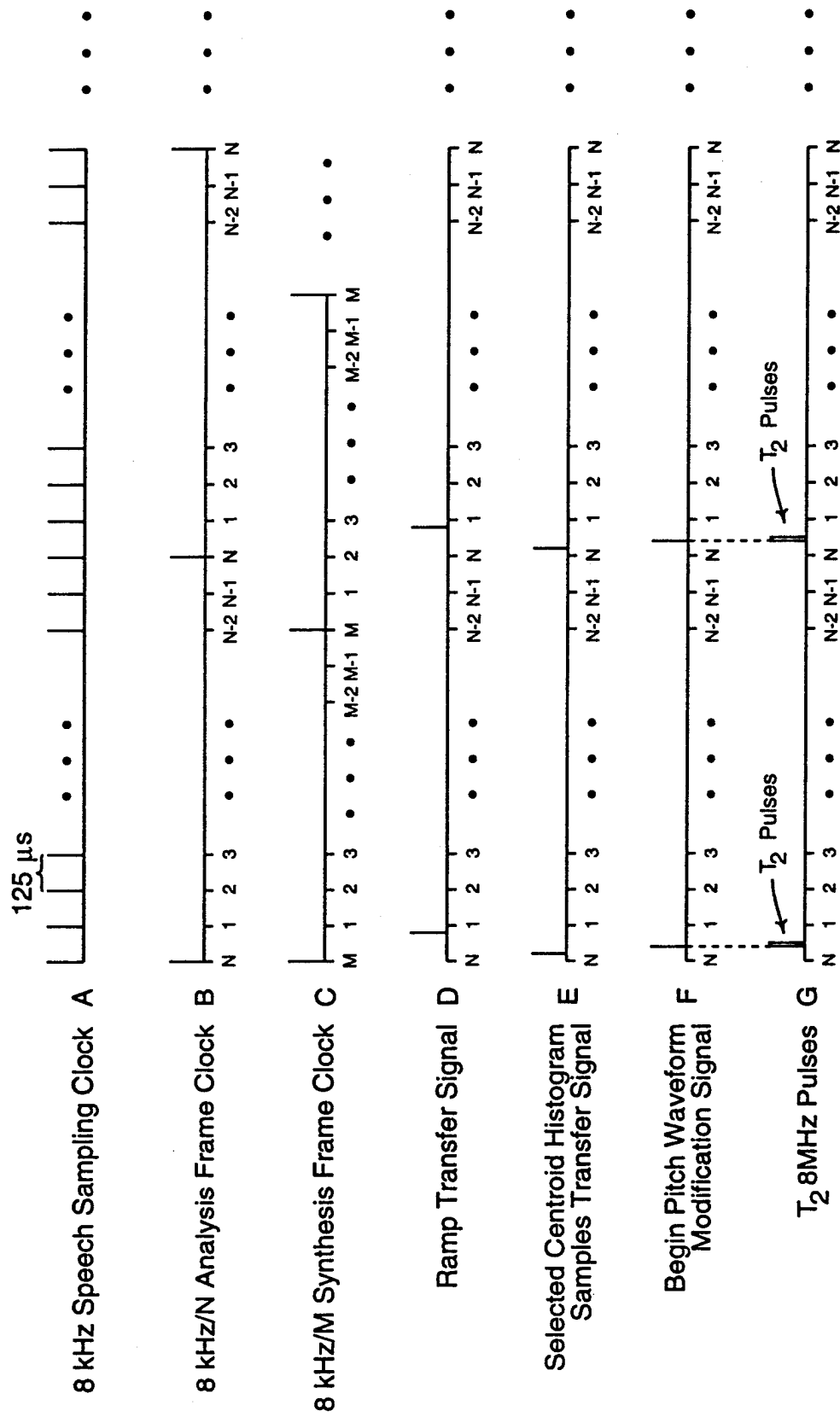


FIG. 10(b)

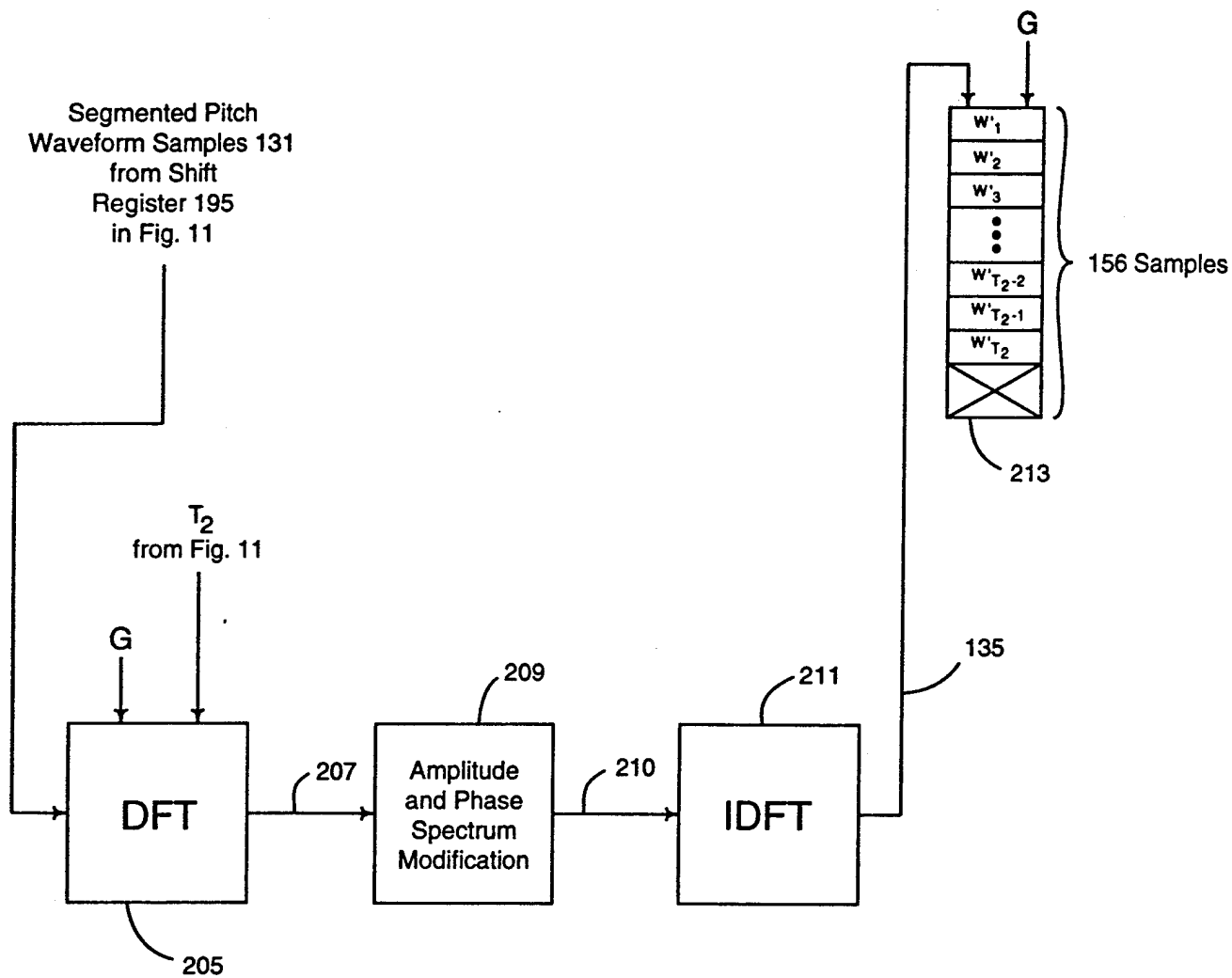


FIG. 12

Segmented Pitch
Waveform Samples 131
from Shift
Register 195
in Fig. 11

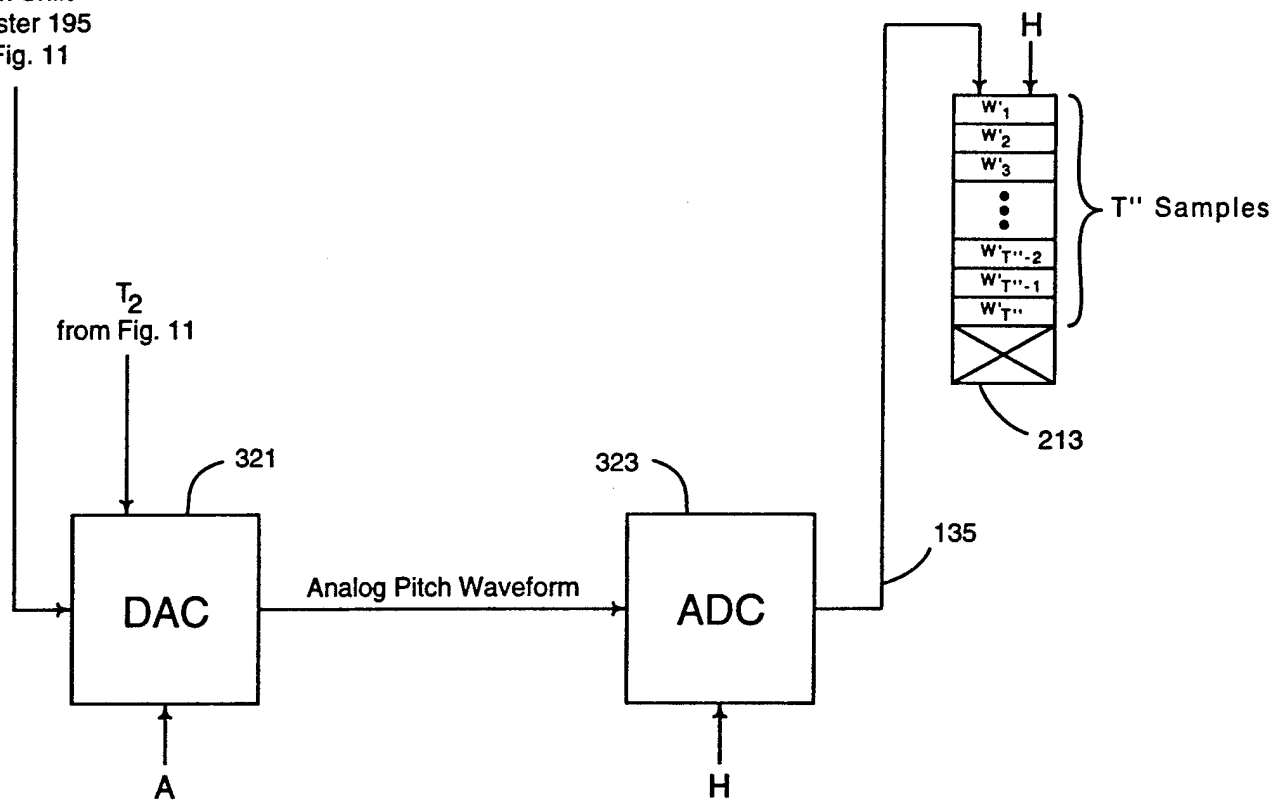


FIG. 13

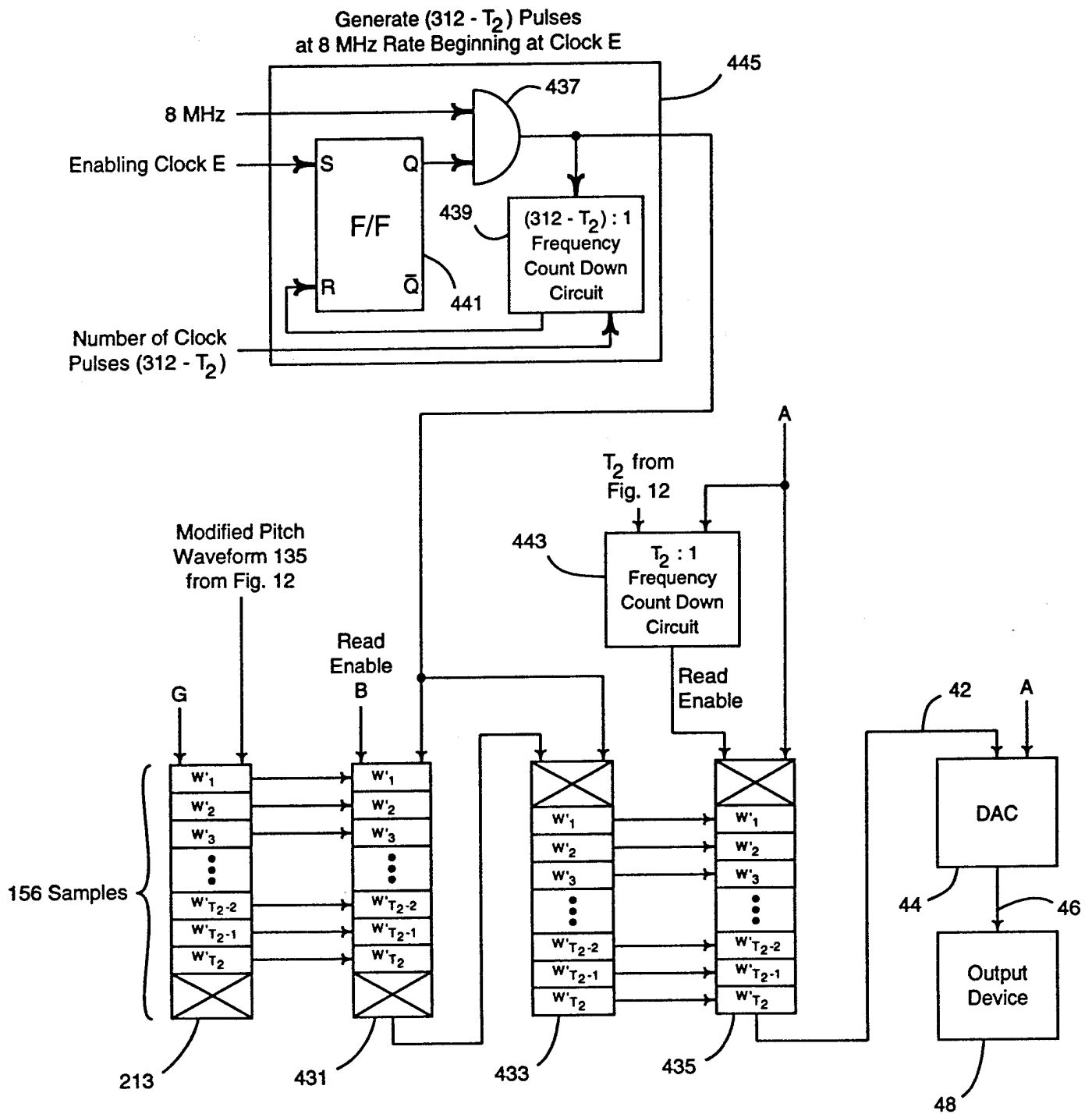


FIG. 14

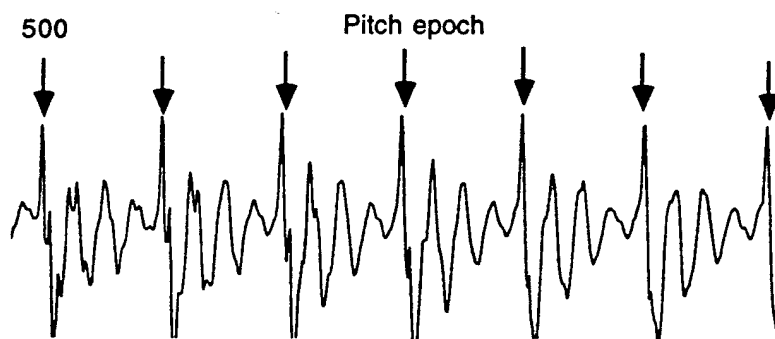


FIG. 15

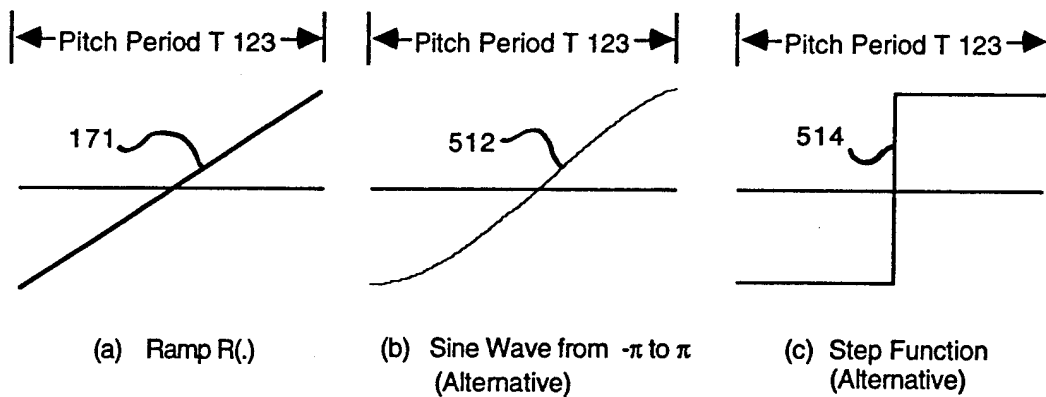


FIG. 16