

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
29 January 2004 (29.01.2004)

PCT

(10) International Publication Number
WO 2004/010627 A1

(51) International Patent Classification⁷: **H04K 1/00**,
G10L 21/02

(21) International Application Number:
PCT/US2003/021578

(22) International Filing Date: 10 July 2003 (10.07.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
10/205,328 24 July 2002 (24.07.2002) US

(71) Applicant: **APPLIED MINDS, INC.** [US/US]; 1209
Grand Central Avenue, Glendale, CA 91201 (US).

(72) Inventors: **HILLIS, W., Daniel**; 10064 Toluca Lake Av-
enue, Toluca Lake, CA 91601 (US). **FERREN, Bran**; 1236

Benedict Canyon Drive, Beverly Hills, CA 90210 (US).
HOWE, Russel; 4343 Ocean View Avenue #179, Mon-
trose, CA 91020 (US). **ENO, Brian**; c/o Opal Ltd., 4 Pem-
bridge Mews, London W11 3EQ (GB).

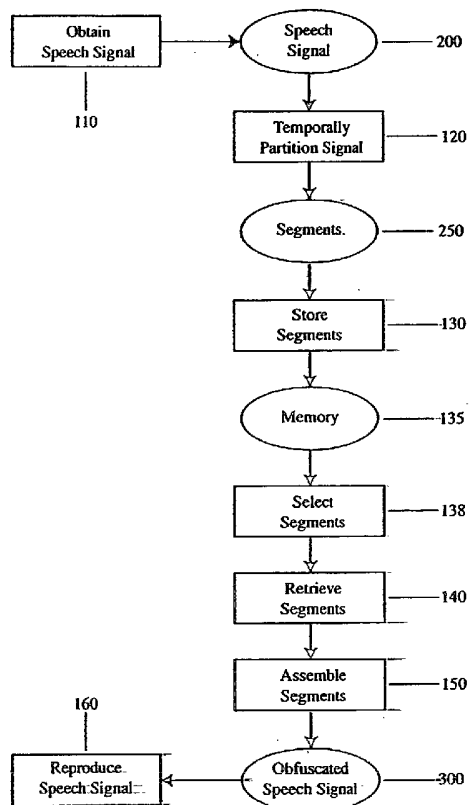
(74) Agents: **GLENN, Michael, A.** et al.; 3475 Edison Way,
Ste. L., Menlo Park, CA 94025 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,
MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE,
SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC,
VN, YU, ZA, ZM, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),
European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE,

[Continued on next page]

(54) Title: METHOD AND SYSTEM FOR MASKING SPEECH



(57) Abstract: A simple and efficient method for producing an obfuscated speech signal which may be used to mask a stream of speech, is disclosed. A speech signal representing the speech stream to be masked is obtained (110). The speech signal (200) is then temporally partitioned (120) into segments (250), preferably corresponding to phonemes within the speech stream. The segments are then stored (130) in a memory (135), and some or all of the segments are subsequently selected (138), retrieved (140), and assembled (150) into an obfuscated speech signal (300) representing an unintelligible speech stream that, when combined with the speech signal or reproduced (160) and combined with the speech stream, provides a masking effect. While the presently preferred embodiment finds application most readily in an open plan office, embodiments suitable for use in restaurants, classrooms, and in telecommunications systems are also disclosed.



ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declaration under Rule 4.17:

- *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW, ARIPO patent (GH, GM,*

KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)

Published:

- *with international search report*
— *with amended claims*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

METHOD AND SYSTEM FOR MASKING SPEECH

BACKGROUND

TECHNICAL FIELD

5

This invention relates to systems for concealing information and, in particular, those systems that render a speech stream unintelligible.

DESCRIPTION OF THE PRIOR ART

10

The human auditory system is very adept at distinguishing and comprehending a stream of speech amid background noise. This ability offers tremendous advantages in most instances because it allows for speech to be understood amid noisy environments.

15

In many instances, though, such as in open plan office spaces, it is highly desirable to mask speech, either to provide privacy to the speaker or to lessen the distraction of those within audible range. In these cases, the human ability to discern speech in the presence of background noise presents special challenges. Simply introducing noise of a stochastic nature, *e.g.* white or pink noise, is typically unsuccessful, in that the amplitude of the introduced noise must be increased to unacceptable levels before the underlying speech can no longer be understood.

25 Accordingly, many prior art approaches to masking speech have focused on generating specialized forms of masking noise, in an effort to lower the

intensity of noise required to render a stream of speech unintelligible. For example, United States Patent 3,985,957 to Torn discloses a "sound masking system" for "masking conversation in an open plan office." In this approach, "a conventional generator of electrical random noise currents feeds its output
5 through adjustable electric filter means to speaker clusters in a plenum above the office space." Despite such sophistication, in many instances the level of background noise required to mask conversation effectively remains unacceptably high.

10 Other approaches have sought to provide masking more discretely by deploying microphones and speakers in more complex physical configurations and controlling them with active noise cancellation algorithms. For example, United States Patent 5,315,661 to Gossman describes a system for
15 "controlling sound transmission through (from) a panel using sensors, actuators and an active control system. The method uses active structural acoustic control to control sound transmission through a number of smaller panel cells which are in turn combined to create a larger panel." It is intended that the invention serve as "a replacement for thick and heavy passive sound isolation material, or anechoic material." While such systems are in theory
20 effective, they are difficult to implement in practice, and are often prohibitively expensive.

Several techniques for performing obfuscation (often termed scrambling) may also be found in the prior art. United States Patent 4,068,094 to Schmid et al.
25 describes "a method of scrambling and unscrambling speech transmissions

by first dividing the speech frequencies into two frequency bands and reversing their order by modulating the speech information.”

Adopting a somewhat different approach, United States Patent 4,099,027 to
5 Whitten discloses a system operating primarily in the time domain. Specifically, “a speech scrambler for rendering unintelligible a communications signal for transmission over nonsecure communications channels includes a time delay modulator and a coding signal generator in a scrambling portion of the system and a similar time delay modulator and a
10 coding generator for generating an inverse signal in the unscrambling portion of the system.”

These methods are effective in producing an obfuscated stream of speech, that when presented in place of the original stream of speech, is unintelligible.
15 However, they are less effective in rendering a stream of speech unintelligible via superposition of the obfuscated stream of speech. This represents a significant deficiency for application to conversation masking in an office environment, where direct substitution of the obfuscated speech stream for the original speech stream is impractical if not impossible. Furthermore, due
20 to the nature of the scrambling, the obfuscated speech stream does not sound speech-like to the listener. In environments such as open plan offices, the obfuscated stream may therefore prove more distracting than the original speech stream.

25 United States Patent 4,195,202 to McCalmont suggests an improvement on these systems that may in fact produce a less intelligible composite stream,

but does not address the need for a speech-like scrambled signal. In fact, a specific effort is made to eliminate one of the key features of human speech. An "encoding apparatus first divides a voice signal to be transmitted into two or more frequency bands. One or more of the frequency bands is frequency
5 inverted, delayed in time relative to the other frequency bands and then recombined with the other frequency bands to produce a composite signal for transmission to a remote receiver. By selecting the magnitude of the delay to approximate the time constants of the cadence, or intersyllabic and phoneme generation rates, of the speech to which the voice signal corresponds, the
10 amplitude fluctuations of the composite signal are substantially lessened and the cadence content of the signal is effectively disguised."

What is needed is a simple and effective system for masking a stream of speech in environments such as open plan offices, where an obfuscated
15 speech stream cannot be substituted for, but merely added to, an original stream of speech. The method should provide an obfuscated speech stream that is speech-like in nature yet highly unintelligible. Furthermore, combination of the original speech stream and obfuscated speech stream should produce a combined speech stream that is also speech-like yet
20 unintelligible.

SUMMARY OF THE INVENTION

The invention provides a simple and efficient method for producing an
25 obfuscated speech signal which may be used to mask a stream of speech. A speech signal representing the speech stream to be masked is obtained. The

speech signal is then temporally partitioned into segments, preferably corresponding to phonemes within the speech stream. The segments are then stored in a memory, and some or all of the segments are subsequently selected, retrieved, and assembled into an obfuscated speech signal
5 representing an unintelligible speech stream that, when combined with the speech signal or reproduced and combined with the speech stream, provides a masking effect.

The obfuscated speech signal may be produced in substantially real time,
10 allowing for direct masking of a speech stream, or may be produced from a recorded speech signal. In creating the obfuscated speech signal, segments within the speech signal may be reordered in a one-to-one fashion, segments may be selected and retrieved at random from a recent history of segments within the speech signal, or segments may be classified or identified and then
15 selected with a relative frequency commensurate with their frequency of occurrence within the speech signal. Finally, it is possible that more than one selection, retrieval, and assembly process may be conducted concurrently to produce more than one obfuscated speech signal.

20 While the presently preferred embodiment of the invention most readily finds application in an open plan office, alternative embodiments may find application, for example, in restaurants, classrooms, and in telecommunications systems.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows a device for masking a speech stream in an open plan office
5 according to the presently preferred embodiment of the invention;

Figure 2 is a flow chart showing a method for producing an obfuscated
speech signal according to the presently preferred embodiment of the
invention;

10

Figure 3 is a detailed flow chart showing a method for temporally partitioning
a speech signal into segments and storing the segments according to the
presently preferred embodiment of the invention; and

15 Figure 4 is a detailed flow chart showing a method for selecting, retrieving,
and assembling segments according to the presently preferred embodiment of
the invention.

20

DESCRIPTION OF THE INVENTION

The invention provides a simple and efficient method for producing an
obfuscated speech signal which may be used to mask a stream of speech.

25 Figure 1 shows a device for masking a speech stream in an open plan office
according to the presently preferred embodiment of the invention. A speaking

office worker 11 in a first cubicle 21 wishes to hold a private conversation. The partition 30 separating the speaking worker's cubicle from an adjacent cubicle 22 does not provide sufficient acoustic isolation to prevent a listening office worker 12 in the adjacent cubicle from overhearing the conversation. This
5 situation is undesirable because the speaking worker is denied privacy and the listening worker is distracted, or worse, may overhear a confidential conversation.

Figure 1 illustrates how the presently preferred embodiment of the invention
10 may be used to remedy this situation. A microphone 40 is placed in a position allowing acquisition of the stream of speech emanating from the speaking worker 11. Preferably, the microphone is mounted in a location where a minimum of acoustic information other than the desired speech stream is captured. A location substantially above the speaking worker 11, but still
15 within the first cubicle 21, may provide satisfactory results.

The signal representing the stream of speech obtained by the microphone is provided to a processor 100 that identifies the phonemes composing the speech stream. In real time or near real time, an obfuscated speech signal is
20 generated from a sequence of phonemes similar to the identified phonemes. When reproduced as an obfuscated speech stream, the obfuscated speech signal is speech-like, yet unintelligible.

The obfuscated speech stream is reproduced and presented, using one or
25 more speakers 50, to those workers who may potentially overhear the speaking worker, including the listening worker 12 in the adjacent cubicle 22.

The obfuscated speech stream, when heard superimposed upon the original speech stream, yields a composite speech stream that is unintelligible, thus masking the original speech stream. Preferably, the obfuscated speech stream is presented at an intensity comparable to that of the original speech stream. Presumably, the listening worker is well accustomed to hearing speech-like sounds emanating from the first cubicle at an intensity commensurate with typical human speech. The listening worker is therefore unlikely to be distracted by the composite speech stream provided by the invention.

10

The speakers 50 are preferably placed in a location where they are audible to the listening worker but not audible to the speaking worker. Additionally, care must be taken to ensure that the listening worker cannot isolate the original speech stream from the obfuscated speech stream using directional cues.

15 Multiple speakers, preferably placed so as not to be coplanar with one another, may be used to create a complex sound field that more effectively masks the original speech stream emanating from the speaking worker. Additionally, the system may use information about the location of the speaker, *e.g.* based upon the location of the microphone, and
20 activate/deactivate various speakers to achieve an optimum dispersion of masking speech. In this regard, an open office environment may be monitored to control speakers and to mix various obfuscated conversations derived from multiple locations so that several conversations may take place, and be masked, simultaneously. For example, the system can direct and weight
25 signals to various speakers based upon information derived from several microphones.

Figure 2 is a flow chart showing a method for producing an obfuscated speech signal according to the presently preferred embodiment of the invention. In the preferred embodiment, this process is conducted by the processor 100 of Figure 1. A speech signal 200 representing the speech stream to be masked is obtained 110 from a microphone or similar source, as shown in Figure 1. The speech signal $s(t)$, is preferably obtained and subsequently manipulated as a discrete series of digital values, $s(n)$. In the preferred embodiment, where the microphone 40 provides an analog signal, this requires that the signal be digitized by an analog-to-digital converter.

Once obtained, the speech signal is temporally partitioned 120 into segments 250. As described above, the segments correspond to phonemes within the speech stream. The segments are then stored 130 in a memory 135, thus allowing selected segments to be subsequently selected 138, retrieved 140, and assembled 150. The result of the assembly operation is an obfuscated speech signal 300 representing an obfuscated speech stream.

The obfuscated speech signal may then be reproduced 160, preferably through one or more speakers as shown in Figure 1. In the preferred embodiment, where the one or more speakers require an analog input signal, this may require the use of a digital- to-analog converter. Alternatively, the speech signal and obfuscated speech signal may be combined, and the combined signal reproduced.

It is important to note that while the flow of data through the above process is as shown in Figure 2, the operations detailed may in practice be executed concurrently, providing substantially steady state processing of data in real time. Alternatively, the process may be conducted as a post-processing operation applied to a pre-recorded speech signal.

Selection 138, retrieval 140, and assembly 150 of the signal segments may be accomplished in any of several manners. In particular, segments within the speech signal may be reordered in a one-to-one fashion, segments may be selected and retrieved at random from a recent history of segments within the speech signal, or segments may be classified or identified and then selected with a relative frequency commensurate with their frequency of occurrence within the speech signal. Furthermore, it is possible that several selection, retrieval, and assembly processes may be conducted concurrently to produce several obfuscated speech signals.

Figure 3 is a detailed flow chart showing a method for temporally partitioning a speech signal into segments and storing the segments according to the presently preferred embodiment of the invention. Here, the steps of temporally partitioning the signal into segments and storing the segments in memory shown in Figure 2 are described in greater detail. The partitioning operation is conducted in a manner such that the resulting segments correspond to phonemes within the speech stream.

To partition the speech signal 200 into segments, the speech signal is squared 122, and the resulting signal $s^2(n)$ is averaged 1231, 1232, 1233 over

three time scales, *i.e.* a short time scale T_s ; a medium time scale T_m ; and a long time scale T_l . The averaging is preferably implemented through the calculation of running estimates of the averages, V_i , according to the expression

5

$$V_i(n+1) = a_i s(n) + (1 - a_i) V_i(n), \quad E[l, m, s]. \quad (1)$$

This is approximately equivalent to a sliding window average of N_i samples, with

10

$$a_i = \frac{1}{N_i} = \frac{1}{f T_i} \quad (2)$$

where f is the sampling rate and T_i the time scale.

15

Preferably, the short time scale T_s is selected to be characteristic of the duration of a typical phoneme and the medium time scale T_m is selected to be characteristic of the duration of a typical word. The long time scale T_l is a conversational time scale, characteristic of the ebb and flow of the speech stream as a whole. In the presently preferred embodiment of the invention, values of 0.125, 0.250, and 1.00 sec, respectively, have provided acceptable system performance, although those skilled in the art will appreciate that this embodiment of the invention may readily be practiced with other time scale values.

25

The result of the medium time scale average 1232 is multiplied 124 by a weighting 125, and then subtracted 126 from the result of the short time scale average 1231. Preferably, the value of the weighting is between 0 and 1, In practice, a value of 1/2 has proven acceptable.

5

The resulting signal is monitored to detect 127 zero crossings. When a zero crossing is detected, a true value is returned. A zero crossing reflects a sudden increase or decrease in the short time scale average of the speech signal energy that could not be tracked by the medium time scale average.

10 Zero crossings thus indicate energy boundaries that generally correspond to phoneme boundaries, providing an indication of the times at which transitions occur between successive phonemes, between a phoneme and a subsequent period of relative silence, or between a period of relative silence and a subsequent phoneme.

15

The result of the long time average 1233 is passed to a threshold operator 128. The threshold operator returns "true" if the long time average is above an upper threshold value and "false" if the long time average is below a lower threshold value. In some embodiments of the invention, the upper and lower
20 threshold values may be the same. In the preferred embodiment, the threshold operator is hysteretic in nature, with differing upper and lower threshold values.

If a speech signal 200 is present and 1292 the threshold operator 128 returns
25 a true value, the speech signal is stored in a buffer 136 within an array of

buffers residing in the memory 135. The particular buffer in which the signal is stored is determined by a storage counter 132.

If a zero crossing is detected 127 and 1291 the threshold operator 128 returns
5 a "true" value, the storage counter 132 is incremented 131, and storage begins in the next buffer 136 within the array of buffers in the memory 135. In this manner, each buffer in the array of buffers is filled with a phoneme or interstitial silence of the speech signal, as partitioned by the detected zero crossings. When the last buffer in the array of buffers is reached, the counter
10 is reset and the contents of the first buffer are replaced with the next phoneme or interstitial silence. Thus, the buffer accumulates and then maintains a recent history of the segments present within the speech signal.

It should be noted that this method represents only one of a variety of ways in
15 which the speech signal may be partitioned into segments corresponding to phonemes. Other algorithms, including those used in continuous speech recognition software packages, may also be employed.

Figure 4 is a detailed flow chart showing a method for selecting, retrieving,
20 and assembling segments according to the presently preferred embodiment of the invention. Here, the steps of selecting 138 segments, retrieving 140 segments from memory and assembling 150 segments into an obfuscated speech signal shown in Figure 2 are presented in greater detail.

25 A random number generator 144 is used to determine the value of a retrieval counter 142. The buffer 136 indicated by the value of the counter is read from

the memory 135. When the end of the buffer is reached, the random number generator provides another value to the retrieval counter, and another buffer is read from memory. The contents of the buffer are appended to the contents of the previously read buffer through a catenation 152 operation to compose
5 the obfuscated speech signal 300. In this manner, a random sequence of signal segments reflecting the recent history of segments within the speech signal 200 are combined to form the obfuscated speech signal 300.

It is often desirable to provide masking only during moments of active
10 conversation. Thus, in the preferred embodiment, buffers are only read from memory if a buffer is available and 139 the threshold operator 128 of Figure 3 returns a "true" value.

Several other noteworthy features have also been incorporated into the
15 presently preferred embodiment of the invention. First, a minimum segment length is enforced. If a zero crossing indicates a phoneme or interstitial silence less than the minimum segment length, the zero crossing is ignored and storage continues in the current buffer 136 within the array of buffers in the memory 135. Also, a maximum phoneme length is enforced, as
20 determined by the size of each buffer in the buffer array. If, during storage, the maximum phoneme length is exceeded, a zero crossing is inferred, and storage begins in the next buffer within the array of buffers. To avoid conflict between storage in and retrieval from the array of buffers, if a particular buffer is currently being read and is simultaneously selected by the storage counter
25 132, the storage counter is again incremented, and storage begins in the next buffer within the array of buffers.

Finally, during the catenation 152 operation, it may be advantageous to apply a shaping function to the head and tail of the segment selected by the retrieval counter 142. The shaping function provides a smoother transition
5 between successive segments in the obfuscated speech signal, thereby yielding a more natural sounding speech stream upon reproduction 160. In the preferred embodiment, each segment is smoothly ramped up at the head of the segment and down at the tail of the segment using a trigonometric function. The ramping is conducted over a time scale shorter than the
10 minimum allowable segment. This smoothing serves to eliminate audible pops, clicks, and ticks at the transitions between successive segments in the obfuscated speech signal.

The masking method described herein may be used in environments other
15 than office spaces. In general, it may be employed anywhere a private conversation may be overheard. Such spaces include, for example, crowded living quarters, public phone booths, and restaurants. The method may also be used in situations where an intelligible stream of speech may be distracting. For example, in open space classrooms, students in one
20 partitioned area may be less distracted by an unintelligible voice-like speech stream emanating from an adjacent area than by a coherent speech stream.

The invention is also easily extended to the emulation of realistic yet unintelligible voice-like background noise. In this application, the modified
25 signal may be generated from a previously obtained voice recording, and presented in an otherwise quiet environment. The resulting sound presents

the illusion that one or more conversations are being conducted nearby. This application would be useful, for example, in a restaurant, where an owner may want to promote the illusion that a relatively empty restaurant is populated by a large number of diners, or in a theatrical production to give the impression of
5 a crowd.

If the specific masking method employed is known to both of two communicating parties, it may be possible to transmit an audio signal secretly using the described technique. In this case, the speech signal
10 would be masked by superposition of the obfuscated speech signal, and unmasked upon reception. It is also possible that the particular algorithm used is seeded by a key known only to the communicating parties, thereby thwarting any attempts by a third party to intercept and unmask the transmission.

15

Although the invention is described herein with reference to the preferred embodiment, one skilled in the art will readily appreciate that other applications may be substituted for those set forth herein without departing from the spirit and scope of the present invention. Accordingly, the invention
20 should only be limited by the Claims included below.

CLAIMS

1. A method of producing a substantially unintelligible, obfuscated speech signal from intelligible speech, comprising the steps of:

- 5 obtaining a speech signal representing a speech stream;
 temporally partitioning said speech signal into a plurality of segments,
said segments occurring in an initial order within said speech signal;
 selecting a plurality of selected segments from among said segments;
and
10 assembling said selected segments, in an order different than said
initial order, to produce said obfuscated speech signal.

2. The method of Claim 1, further comprising the step, immediately following said temporally partitioning step, of:

- 15 storing said segments in a memory; and
 further comprising the step, immediately following said selecting step,
of:
 retrieving said selected segments from said memory.

20 3. The method of Claim 1, wherein said obfuscated speech signal is produced in substantially real time.

4. The method of Claim 1, wherein said speech signal represents a previously recorded speech stream.

25

5. The method of Claim 1, wherein said obfuscated speech signal simulates unintelligible background conversation.

6. The method of Claim 1, wherein said obfuscated speech signal is
5 transmitted through a telecommunications network.

7. The method of Claim 1, further comprising the step, immediately following said assembling step, of:

combining said speech signal and said obfuscated speech signal to
10 produce a combined speech signal;

wherein said combined signal comprises a speech stream that is substantially unintelligible.

8. The method of Claim 1, further comprising the steps, immediately following
15 said assembling step, of:

reproducing said obfuscated speech signal to provide an obfuscated speech stream, and

combining said speech stream and said obfuscated speech stream to produce a combined speech stream;

20 wherein said combined speech stream is substantially unintelligible.

9. The method of Claim 1, wherein said speech signal is obtained from a microphone.

25 10. The method of Claim 1, wherein said obfuscated speech signal is reproduced by a loudspeaker.

11. The method of Claim 1, wherein said speech signal is obtained from an office environment.

12. The method of Claim 1, wherein said selected segments comprise each
5 segment within said speech stream.

13. The method of Claim 2, wherein said selected segments are selected from a plurality of segments within said memory comprising a recent history of segments present in said speech signal.

10

14. The method of Claim 13, wherein said selected segments are selected randomly from said plurality of segments contained within said memory.

15. The method of Claim 13, wherein each of said selected segments is
15 selected with a relative frequency commensurate with a relative frequency of occurrence within said speech signal.

16. The method of Claim 1, wherein said speech signal comprises a sequence of digital values.

20

17. The method of Claim 1, wherein said segments represent phonemes within said speech stream.

18. The method of Claim 17, wherein said phonemes are determined using a
25 continuous speech recognition system.

19. The method of Claim 17, wherein said temporally partitioning step comprises the steps of:

squaring said speech signal;

calculating a short time average of said speech signal over a short time

5 scale;

calculating a medium time average of said speech signal over a medium time scale;

calculating a difference between said short time average and said medium time average; and

10 detecting zero crossings in said difference;

wherein said zero crossings delineate said segments.

20. The method of Claim 19, wherein said short time scale characterizes a length of a typical phoneme in said speech stream.

15

21. The method of Claim 19, wherein said medium time scale characterizes a length of a typical word in said speech stream.

22. The method of Claim 2, wherein said storing step comprises the steps of:

20 squaring said speech signal;

calculating a long time average of said speech signal over a long time scale;

determining when said long time average is above a first threshold and when said long time average is below a second threshold;

25 halting said storing of said segments in said memory when said long time average is below said second threshold; and

resuming said storing of said segments in said memory when said long time average is above said first threshold.

23. The method of Claim 22, wherein said long time scale characterizes a conversational time scale of said speech stream.

24. The method of Claim 2, wherein said retrieving step comprises the steps of:

squaring said speech signal;

10 calculating a long time average of said speech signal over a long time scale;

determining when said long time average is above a first threshold and when said long time average is below a second threshold;

halting said retrieving of said segments from said memory when said

15 long time average is below said second threshold; and

resuming said retrieving of said segments from said memory when said long time average is above said first threshold.

25. The method of Claim 24, wherein said long time scale characterizes a conversational time scale of said speech stream.

26. The method of Claim 1, wherein said assembling step comprises the step of:

applying a shaping function to each of said selected segments;

25 wherein said shaping function provides a smooth transition between successive segments in said obfuscated speech signal.

27. The method of Claim 1, wherein said selecting and assembling steps concurrently produce a plurality of said obfuscated speech signals from said speech signal.

5

28. A method of masking a speech stream, comprising the steps of:

obtaining a speech signal representing said speech stream;

modifying said speech signal to create an obfuscated speech signal;

and

10 combining said speech signal and said obfuscated speech signal to produce a combined speech signal;

wherein said combined speech signal represents a combined speech stream that is substantially unintelligible.

15 29. A method of masking a speech stream, comprising the steps of:

obtaining a speech signal representing said speech stream;

modifying said speech signal to create an obfuscated speech signal;

reproducing said obfuscated speech signal to provide an obfuscated speech stream; and

20 combining said speech stream and said obfuscated speech stream to produce a combined speech stream;

wherein said combined speech stream is substantially unintelligible.

30. An apparatus for producing a substantially unintelligible, obfuscated

25 speech signal from intelligible speech, comprising:

a module for obtaining a speech signal representing a speech stream;

a module for temporally partitioning said speech signal into a plurality of segments, said segments occurring in an initial order within said speech signal;

a module for selecting a plurality of selected segments from among
5 said segments; and

a module for assembling said selected segments, in an order different than said initial order, to produce said obfuscated speech signal.

31. The apparatus of Claim 30, further comprising:

10 a memory for storing said segments; and

a module for retrieving said selected segments from said memory.

32. The apparatus of Claim 30, wherein said obfuscated speech signal is produced in substantially real time.

15

33. The apparatus of Claim 30, wherein said speech signal represents a previously recorded speech stream.

34. The apparatus of Claim 30, wherein said obfuscated speech signal
20 simulates unintelligible background conversation.

35. The apparatus of Claim 30, further comprising:

a module for transmitting said obfuscated speech signal through a telecommunications network.

25

36. The apparatus of Claim 30, further comprising:

a module for combining said speech signal and said obfuscated speech signal to produce a combined speech signal;

wherein said combined signal comprises a speech stream that is substantially unintelligible.

5

37. The apparatus of Claim 30, further comprising:

a module for reproducing said obfuscated speech signal to provide an obfuscated speech stream, and

a module for combining said speech stream and said obfuscated speech stream to produce a combined speech stream;

10

wherein said combined speech stream is substantially unintelligible.

38. The apparatus of Claim 30, further comprising:

a microphone for obtaining said speech signal.

15

39. The apparatus of Claim 30, further comprising:

a loudspeaker for reproducing said obfuscated speech.

40. The apparatus of Claim 30, wherein said speech signal is obtained from

an office environment.

20

41. The apparatus of Claim 30, wherein said selected segments comprise each segment within said speech stream.

42. The apparatus of Claim 31, wherein said selected segments are selected

from a plurality of segments within said memory comprising a recent history of segments present in said speech signal.

25

43. The apparatus of Claim 42, wherein said selected segments are selected randomly from said plurality of segments contained within said memory.

5 44. The apparatus of Claim 42, wherein each of said selected segments is selected with a relative frequency commensurate with a relative frequency of occurrence within said speech signal.

45. The apparatus of Claim 30, wherein said speech signal comprises a
10 sequence of digital values.

46. The apparatus of Claim 30, wherein said segments represent phonemes within said speech stream.

15 47. The apparatus of Claim 46, wherein said phonemes are determined using a continuous speech recognition system.

48. The apparatus of Claim 30, wherein said module for temporally partitioning further comprises:

20 a module for squaring said speech signal;
 a module for calculating a short time average of said speech signal over a short time scale;
 a module for calculating a medium time average of said speech signal over a medium time scale;
25 a module for calculating a difference between said short time average and said medium time average; and

a module for detecting zero crossings in said difference;
wherein said zero crossings delineate said segments.

49. The apparatus of Claim 48, wherein said short time scale characterizes a
5 length of a typical phoneme in said speech stream.

50. The apparatus of Claim 48, wherein said medium time scale characterizes
a length of a typical word in said speech stream.

10 51. The apparatus of Claim 31, wherein said memory further comprises:
a module for squaring said speech signal;
a module for calculating a long time average of said speech signal over
a long time scale;
a module for determining when said long time average is above a first
15 threshold and when said long time average is below a second threshold;
a module for halting said storing of said segments in said memory
when said long time average is below said second threshold; and
a module for resuming said storing of said segments in said memory
when said long time average is above said first threshold.

20

52. The apparatus of Claim 51, wherein said long time scale characterizes a
conversational time scale of said speech stream.

53. The apparatus of Claim 31, wherein said module for retrieving:
25 a module for squaring said speech signal;

a module for calculating a long time average of said speech signal over a long time scale;

a module for determining when said long time average is above a first threshold and when said long time average is below a second threshold;

5 a module for halting said retrieving of said segments from said memory when said long time average is below said second threshold; and

a module for resuming said retrieving of said segments from said memory when said long time average is above said first threshold.

10 54. The apparatus of Claim 53, wherein said long time scale characterizes a conversational time scale of said speech stream.

55. The apparatus of Claim 30, wherein said module for assembling further comprises:

15 a module for applying a shaping function to each of said selected segments;

wherein said shaping function provides a smooth transition between successive segments in said obfuscated speech signal.

20 56. The apparatus of Claim 30, wherein said modules for selecting and assembling concurrently produce a plurality of said obfuscated speech signals from said speech signal.

57. An apparatus for masking a speech stream, comprising:

25 a module for obtaining a speech signal representing said speech stream;

a module for modifying said speech signal to create an obfuscated speech signal; and

a module for combining said speech signal and said obfuscated speech signal to produce a combined speech signal;

5 wherein said combined speech signal represents a combined speech stream that is substantially unintelligible.

58. An apparatus of masking a speech stream, comprising:

10 a module for obtaining a speech signal representing said speech stream;

a module for modifying said speech signal to create an obfuscated speech signal;

a module for reproducing said obfuscated speech signal to provide an obfuscated speech stream; and

15 a module for combining said speech stream and said obfuscated speech stream to produce a combined speech stream;

wherein said combined speech stream is substantially unintelligible.

PCT/US03/21578

AMENDED CLAIMS

[Received by the International Bureau on 05 December 2003 (05.12.03)
original claims 1-58 replaced by new claims 1-56]

REPLACEMENT CLAIMS

1. A method of producing a substantially unintelligible, obfuscated speech signal from intelligible speech, comprising the steps of:

- 5 obtaining a speech signal representing a speech stream;
 temporally partitioning said speech signal into a plurality of variable segments,
said segments occurring in an initial order within said speech signal;
 storing said segments in a memory;
 selecting a number of segments from among said stored segments;
10 retrieving said selected segments from said memory; and
 assembling said retrieved segments, in an order different than said initial
order, to produce an obfuscated speech signal.

- 15 2. The method of Claim 1, wherein said obfuscated speech signal is produced in
substantially real time.

3. The method of Claim 1, wherein said speech signal represents a previously recorded speech stream.

- 20 4. The method of Claim 1, wherein said obfuscated speech signal simulates
unintelligible background conversation.

5. The method of Claim 1, wherein said obfuscated speech signal is transmitted through a telecommunications network.

25

6. The method of Claim 1, further comprising the step, immediately following said assembling step, of:

combining said speech signal and said obfuscated speech signal to produce a combined speech signal;

5 wherein said combined signal comprises a speech stream that is substantially unintelligible.

7. The method of Claim 1, further comprising the steps, immediately following said assembling step, of:

10 reproducing said obfuscated speech signal to provide an obfuscated speech stream, and

combining said speech stream and said obfuscated speech stream to produce a combined speech stream;

wherein said combined speech stream is substantially unintelligible.

15

8. The method of Claim 1, wherein said speech signal is obtained from a microphone.

9. The method of Claim 1, wherein said obfuscated speech signal is reproduced by
20 a loudspeaker.

10. The method of Claim 1, wherein said speech signal is obtained from an office environment.

25 11. The method of Claim 1, wherein said selected segments comprise each segment within said speech stream.

12. The method of Claim 1, wherein said selected segments are selected from a plurality of segments within said memory comprising a recent history of segments present in said speech signal.

5

13. The method of Claim 12, wherein said selected segments are selected randomly from said plurality of segments contained within said memory.

14. The method of Claim 12, wherein each of said selected segments is selected
10 with a relative frequency commensurate with a relative frequency of occurrence within said speech signal.

15. The method of Claim 1, wherein said speech signal comprises a sequence of digital values.

15

16. The method of Claim 1, wherein said plurality of segments represent phonemes within said speech stream.

17. The method of Claim 16, wherein said phonemes are determined using a
20 continuous speech recognition system.

18. The method of Claim 16, wherein said temporally partitioning step comprises the steps of:

squaring said speech signal;

25 calculating a short time average of said speech signal over a short time scale;

calculating a medium time average of said speech signal over a medium time scale;

calculating a difference between said short time average and said medium time average; and

5 detecting zero crossings in said difference;

wherein said zero crossings delineate said segments.

19. The method of Claim 18, wherein said short time scale characterizes a length of a typical phoneme in said speech stream.

10

20. The method of Claim 18, wherein said medium time scale characterizes a length of a typical word in said speech stream.

21. The method of Claim 1, wherein said storing step comprises the steps of:

15 squaring said speech signal;

calculating a long time average of said speech signal over a long time scale;

determining when said long time average is above a first threshold and when said long time average is below a second threshold;

20 halting said storing of said segments in said memory when said long time average is below said second threshold; and

resuming said storing of said segments in said memory when said long time average is above said first threshold.

22. The method of Claim 21, wherein said long time scale characterizes a conversational time scale of said speech stream.

25

23. The method of Claim 1, wherein said retrieving step comprises the steps of:

squaring said speech signal;

calculating a long time average of said speech signal over a long time scale;

determining when said long time average is above a first threshold and when

5 said long time average is below a second threshold;

halting said retrieving of said segments from said memory when said long time average is below said second threshold; and

resuming said retrieving of said segments from said memory when said long time average is above said first threshold.

10

24. The method of Claim 23, wherein said long time scale characterizes a conversational time scale of said speech stream.

25. The method of Claim 1, wherein said assembling step comprises the step of:

15 applying a shaping function to each of said selected segments;

wherein said shaping function provides a smooth transition between successive segments in said obfuscated speech signal.

26. The method of Claim 1, wherein said selecting and assembling steps

20 concurrently produce a plurality of obfuscated speech signals from said speech signal.

27. A method of masking a speech stream, comprising the steps of:

obtaining a speech signal representing said speech stream;

25 modifying said speech signal to create an obfuscated speech signal; and

combining said speech signal and said obfuscated speech signal to produce a combined speech signal;

wherein said combined speech signal represents a combined speech stream that is substantially unintelligible.

5

28. A method of masking a speech stream, comprising the steps of:

obtaining a speech signal representing said speech stream;

modifying said speech signal to create an obfuscated speech signal;

reproducing said obfuscated speech signal to provide an obfuscated speech

10 stream; and

combining said speech stream and said obfuscated speech stream to produce a combined speech stream;

wherein said combined speech stream is substantially unintelligible.

15 29. An apparatus for producing a substantially unintelligible, obfuscated speech signal from intelligible speech, comprising:

a module for obtaining a speech signal representing a speech stream;

a module for temporally partitioning said speech signal into a plurality of variable segments, said segments occurring in an initial order within said speech

20 signal;

a memory for storing said plurality of segments;

a module for selecting a number of segments from among said stored segments;

a module for retrieving said selected segments from said memory; and

25 a module for assembling said retrieved segments, in an order different than said initial order, to produce an obfuscated speech signal.

30. The apparatus of Claim 29, wherein said obfuscated speech signal is produced in substantially real time.

5 31. The apparatus of Claim 29, wherein said speech signal represents a previously recorded speech stream.

32. The apparatus of Claim 29, wherein said obfuscated speech signal simulates unintelligible background conversation.

10

33. The apparatus of Claim 29, further comprising:

a module for transmitting said obfuscated speech signal through a telecommunications network.

15 34. The apparatus of Claim 29, further comprising:

a module for combining said speech signal and said obfuscated speech signal to produce a combined speech signal;

wherein said combined signal comprises a speech stream that is substantially unintelligible.

20

35. The apparatus of Claim 29, further comprising:

a module for reproducing said obfuscated speech signal to provide an obfuscated speech stream, and

a module for combining said speech stream and said obfuscated speech stream to produce a combined speech stream;

25

wherein said combined speech stream is substantially unintelligible.

36. The apparatus of Claim 29, further comprising:

a microphone for obtaining said speech signal.

5 37. The apparatus of Claim 29, further comprising:

a loudspeaker for reproducing said obfuscated speech.

38. The apparatus of Claim 29, wherein said speech signal is obtained from an office environment.

10

39. The apparatus of Claim 29, wherein said selected segments comprise each segment within said speech stream.

15

40. The apparatus of Claim 29, wherein said selected segments are selected from a plurality of segments within said memory comprising a recent history of segments present in said speech signal.

41. The apparatus of Claim 40, wherein said selected segments are selected randomly from said plurality of segments contained within said memory.

20

42. The apparatus of Claim 40, wherein each of said selected segments is selected with a relative frequency commensurate with a relative frequency of occurrence within said speech signal.

25

43. The apparatus of Claim 29, wherein said speech signal comprises a sequence of digital values.

44. The apparatus of Claim 29, wherein said segments represent phonemes within said speech stream.

5 45. The apparatus of Claim 44, wherein said phonemes are determined using a continuous speech recognition system.

46. The apparatus of Claim 29, wherein said module for temporally partitioning further comprises:

- 10 a module for squaring said speech signal;
 a module for calculating a short time average of said speech signal over a short time scale;
 a module for calculating a medium time average of said speech signal over a medium time scale;
- 15 a module for calculating a difference between said short time average and said medium time average; and
 a module for detecting zero crossings in said difference;
 wherein said zero crossings delineate said segments.

20 47. The apparatus of Claim 46, wherein said short time scale characterizes a length of a typical phoneme in said speech stream.

48. The apparatus of Claim 46, wherein said medium time scale characterizes a length of a typical word in said speech stream.

25

49. The apparatus of Claim 29, wherein said memory further comprises:

a module for squaring said speech signal;
a module for calculating a long time average of said speech signal over a long time scale;
a module for determining when said long time average is above a first threshold and when said long time average is below a second threshold;
5 a module for halting said storing of said segments in said memory when said long time average is below said second threshold; and
a module for resuming said storing of said segments in said memory when said long time average is above said first threshold.

10

50. The apparatus of Claim 49, wherein said long time scale characterizes a conversational time scale of said speech stream.

51. The apparatus of Claim 29, wherein said module for retrieving:

15

a module for squaring said speech signal;
a module for calculating a long time average of said speech signal over a long time scale;
a module for determining when said long time average is above a first threshold and when said long time average is below a second threshold;
20 a module for halting said retrieving of said segments from said memory when said long time average is below said second threshold; and
a module for resuming said retrieving of said segments from said memory when said long time average is above said first threshold.

25

52. The apparatus of Claim 51, wherein said long time scale characterizes a conversational time scale of said speech stream.

53. The apparatus of Claim 29, wherein said module for assembling further comprises:

a module for applying a shaping function to each of said selected segments;

5 wherein said shaping function provides a smooth transition between successive segments in said obfuscated speech signal.

54. The apparatus of Claim 29, wherein said modules for selecting and assembling concurrently produce a plurality of obfuscated speech signals from said speech
10 signal.

55. An apparatus for masking a speech stream, comprising:

a module for obtaining a speech signal representing said speech stream;

a module for modifying said speech signal to create an obfuscated speech
15 signal; and

a module for combining said speech signal and said obfuscated speech signal to produce a combined speech signal;

wherein said combined speech signal represents a combined speech stream that is substantially unintelligible.

20

56. An apparatus of masking a speech stream, comprising:

a module for obtaining a speech signal representing said speech stream;

a module for modifying said speech signal to create an obfuscated speech
signal;

25 a module for reproducing said obfuscated speech signal to provide an obfuscated speech stream; and

a module for combining said speech stream and said obfuscated speech stream to produce a combined speech stream;
wherein said combined speech stream is substantially unintelligible.

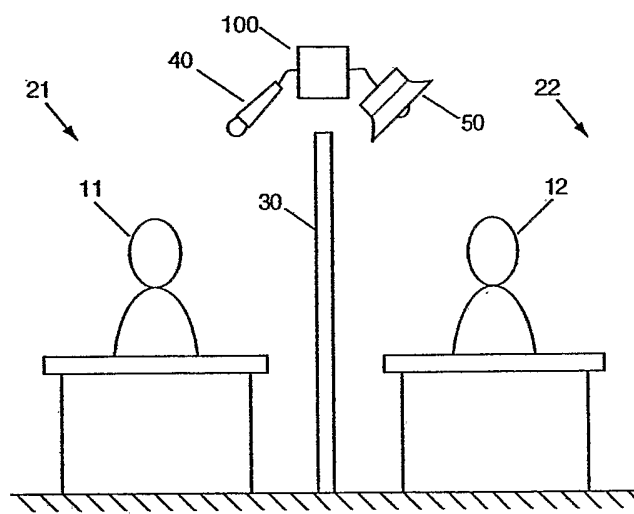


Figure 1

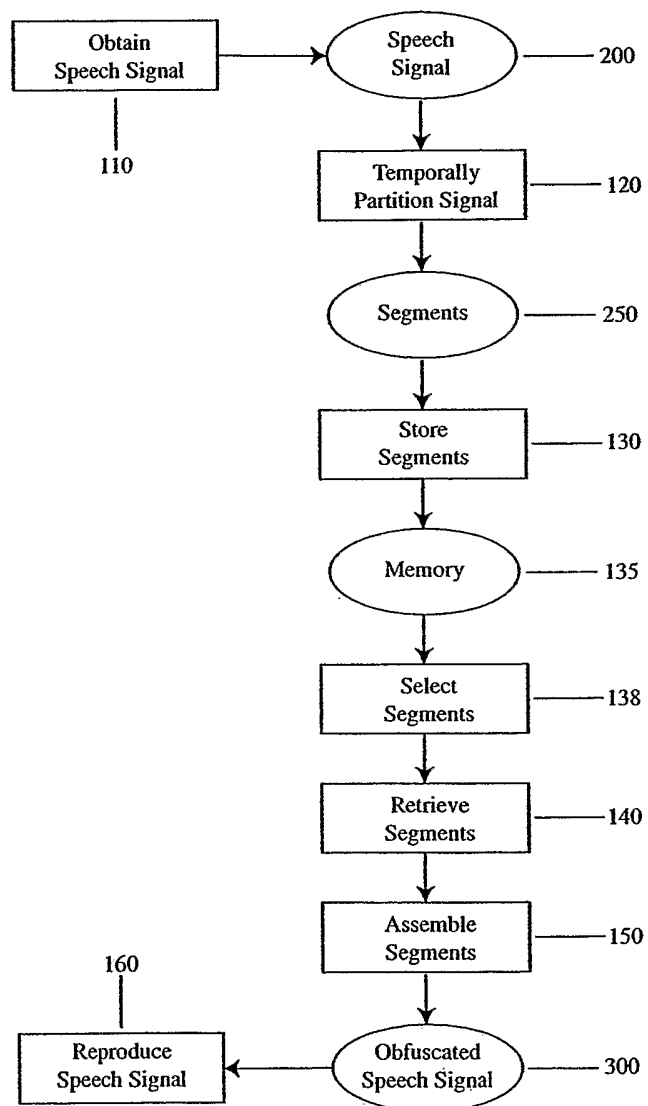


Figure 2

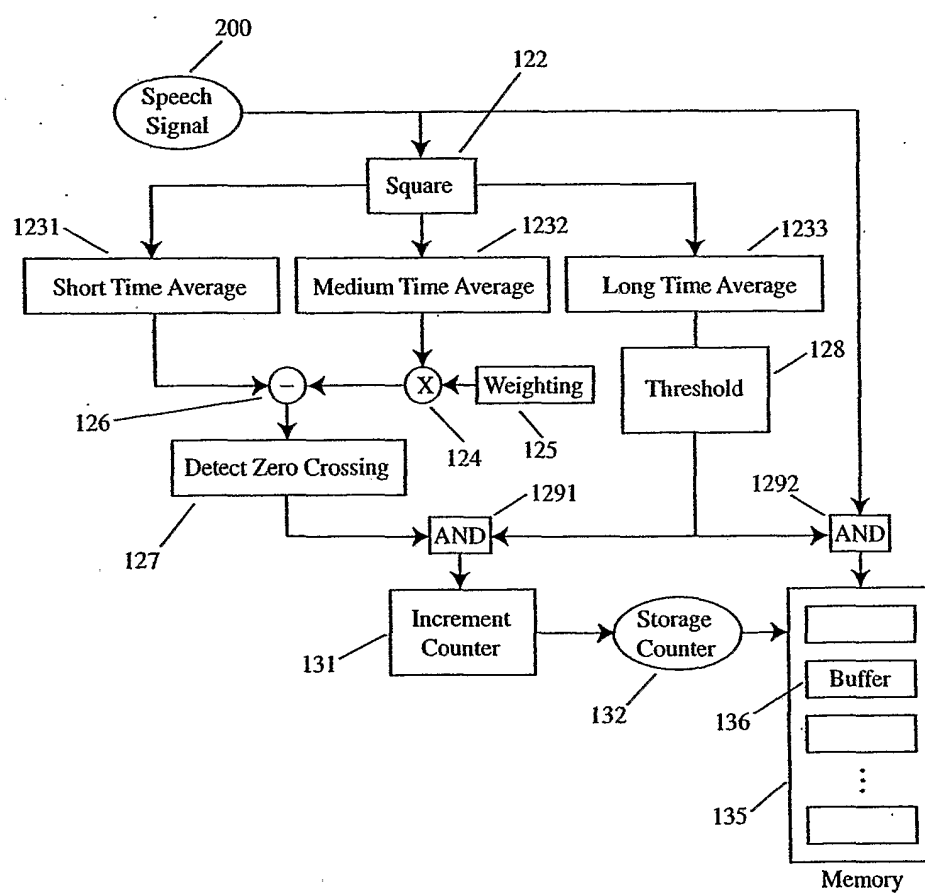


Figure 3

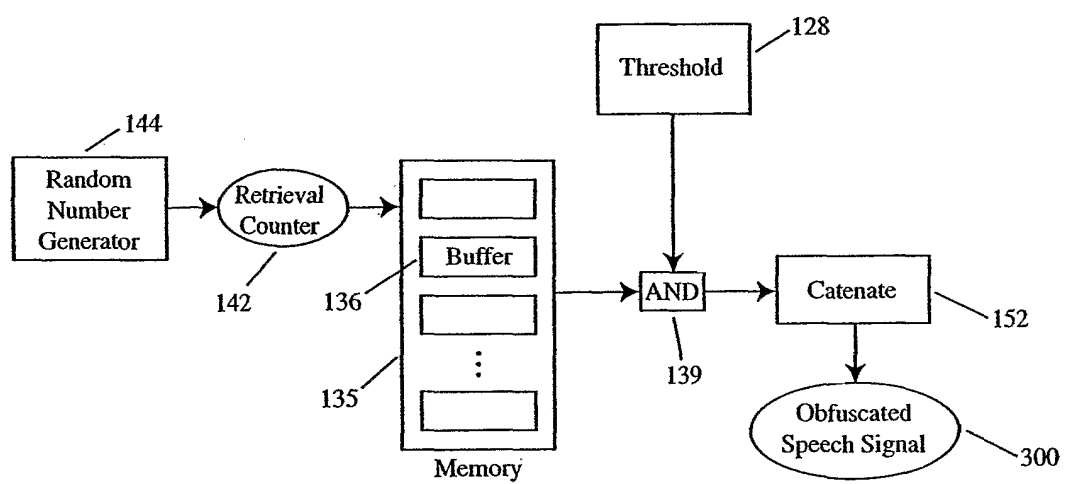


Figure 4

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US03/21578

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : H04K 1/00; G10L 21/02
US CL : 380/36, 275; 704/211, 226, 278

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 704/211, 226, 227, 228, 233, 270, 273, 278; 380/35, 36, 275, 276

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

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Please See Continuation Sheet

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	US 4,905,278 A (PARKER) 27 February 1990 (27.02.1990), column 3, line 15 to column 4, line 11; column 6, lines 40 to 67; column 2, lines 11 to 19: Figure 2.	1-2,4-5,9-10,12-16,26,30-31,33-34,38-39,41-45,55 ----- 3,6-8,11,17-18,27,32,35-37,40,46-47,56
X --- Y	US 4,068,094 A (SCHMID et al.) 10 January 1978 (10.01.1978), column 8, lines 11 to 66: Figure 8.	28,29,57,58 ----- 6-8,11,27,35-37,40,56
Y	US 4,852,170 A (BORDEAUX) 25 July 1989 (25.07.1989), column 2, lines 44 to 59; column 2, lines 64 to 67; column 2, lines 1 to 4.	3,17-18,32,46-47
A	US 4,195,202 A (MCCALMONT) 25 March 1980 (25.03.1980), column 6, lines 37 to 56.	1-58
A	US 3,718,765 A (HALABY) 27 February 1973 (27.02.1973), column 2, lines 1 to 41.	1-58



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:		"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier application or patent published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search

25 August 2003 (25.08.2003)

Date of mailing of the international search report

06 OCT 2003

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US
Commissioner for Patents
P.O. Box 1450
Alexandria, Virginia 22313-1450

Facsimile No. (703)305-3230

Authorized officer

Richemond Dorvil

Telephone No. (703) 305-4000

Eugenia Zagan

INTERNATIONAL SEARCH REPORT

PCT/US03/21578

C. (Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 4,099,027 A (WHITTEN) 04 July 1978 (04.07.1978), column 1, line 2 to column 3, line 33.	1-58
A	US 4,232,194 A (ADAMS) 04 November 1980 (04.11.1980), column 2, lines 11 to 23.	1-58

INTERNATIONAL SEARCH REPORT

PCT/US03/21578

Continuation of B. FIELDS SEARCHED Item 3:

EAST search terms: speech and unintelligible; speech and scrambl\$; phoneme adj boundary; and encrypt\$; phoneme adj duration same (milliseconds or ms); and speech adj recognition