



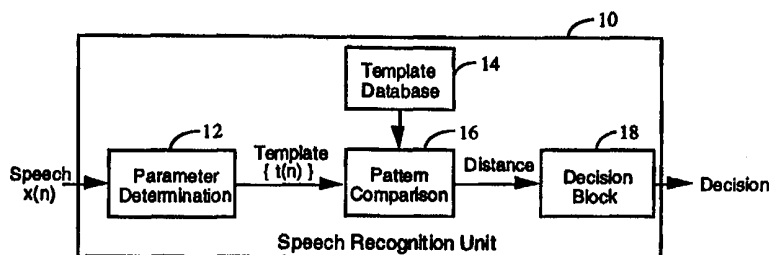
INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : G10L 5/06		A1	(11) International Publication Number: WO 99/40571
			(43) International Publication Date: 12 August 1999 (12.08.99)
(21) International Application Number: PCT/US99/02280 (22) International Filing Date: 3 February 1999 (03.02.99) (30) Priority Data: 09/018,257 4 February 1998 (04.02.98) US (71) Applicant: QUALCOMM INCORPORATED [US/US]; 6455 Lusk Boulevard, San Diego, CA 92121 (US). (72) Inventors: SIH, Gilbert, C.; 7804 Pipat Place, San Diego, CA 92129 (US). BI, Ning; 8370 Capricorn Way #1, San Diego, CA 92126 (US). (74) Agents: OGROD, Gregory, D. et al.; Qualcomm Incorporated, 6455 Lusk Boulevard, San Diego, CA 92121 (US).			(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, 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, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG). Published <i>With international search report.</i>

(54) Title: SYSTEM AND METHOD FOR NOISE-COMPENSATED SPEECH RECOGNITION

(57) Abstract

A system and method for improving speech recognition accuracy in the presence of noise is provided. The speech recognition training unit is modified to store digitized speech samples into a speech database that can be accessed at recognition time. The improved recognition unit comprises a noise analysis, modelling, and synthesis unit which continually analyzes the noise characteristics present in the audio environment and produces an estimated noise signal with similar characteristics. The recognition unit then constructs a noise-compensated template database by adding the estimated noise signal to each of the speech samples in the speech database and performing parameter determination on the resulting sums. This procedure accounts for the presence of noise in the recognition phase by retraining all the templates using an estimated noise signal with similar characteristics as the actual noise signal that corrupted the word to be recognized. This method improves the likelihood of a good template match, which increases the recognition accuracy.



FOR THE PURPOSES OF INFORMATION ONLY

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

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakhstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

SYSTEM AND METHOD FOR NOISE-COMPENSATED SPEECH RECOGNITION

BACKGROUND OF THE INVENTION

5

I. Field of the Invention

The present invention relates to speech processing. More particularly,
the present invention relates to a system and method for the automatic
10 recognition of spoken words or phrases.

II. Description of the Related Art

Digital processing of speech signals has found widespread use,
15 particularly in cellular telephone and PCS applications. One digital speech
processing technique is that of speech recognition. The use of speech
recognition is gaining importance due to safety reasons. For example,
speech recognition may be used to replace the manual task of pushing
buttons on a cellular phone keypad. This is especially important when a
20 user is initiating a telephone call while driving a car. When using a phone
without speech recognition, the driver must remove one hand from the
steering wheel and look at the phone keypad while pushing the buttons to
dial the call. These acts increase the likelihood of a car accident. Speech
recognition allows the driver to place telephone calls while continuously
25 watching the road and maintaining both hands on the steering wheel.
Handsfree carkits containing speech recognition will likely be a legislated
requirement in future systems for safety reasons.

Speaker-dependent speech recognition, the most common type in use
today, operates in two phases: a training phase and a recognition phase. In
30 the training phase, the speech recognition system prompts the user to speak
each of the words in the vocabulary once or twice so it can learn the
characteristics of the user's speech for these particular words or phrases.
The recognition vocabulary sizes are typically small (less than 50 words) and
the speech recognition system will only achieve high recognition accuracy
35 on the user that trained it. An example of a vocabulary for a handsfree
carkit system would include the digits on the keypad, the keywords "call",
"send", "dial", "cancel", "clear", "add", "delete", "history", "program", "yes",
and "no", as well as 20 names of commonly-called coworkers, friends, or
family members. Once training is complete, the user can initiate calls in the
40 recognition phase by speaking the trained keywords. For example, if the

name "John" was one of the trained names, the user can initiate a call to John by saying the phrase "Call John." The speech recognition system recognizes the words "Call" and "John", and dials the number that the user had previously entered as John's telephone number.

5 A block diagram of a training unit 6 of a speaker-dependent speech recognition system is shown in FIG. 1. Training unit 6 receives as input $s(n)$, a set of digitized speech samples for the word or phrase to be trained. The speech signal $s(n)$ is passed through parameter determination block 7, which produces a template of N parameters $\{ p(n) \ n=1...N \}$ capturing the
10 characteristics of the user's pronunciation of the particular word or phrase. Parameter determination unit 7 may implement any of a number of speech parameter determination techniques, many of which are well-known in the art. An exemplary embodiment of a parameter determination technique is the vocoder encoder described in U.S. Pat. No. 5,414,796, entitled
15 "VARIABLE RATE VOCODER," which is assigned to the assignee of the present invention and incorporated by reference herein. An alternative embodiment of a parameter determination technique is a fast fourier transform (FFT), where the N parameters are the N FFT coefficients. Other embodiments derive parameters based on the FFT coefficients. Each spoken
20 word or phrase produces one template of N parameters that is stored in template database 8. After training is completed over M vocabulary words, template database 8 contains M templates, each containing N parameters. Template database 8 is stored into some type of non-volatile memory so that the templates stay resident when the power is turned off.

25 FIG. 2 is a block diagram of speech recognition unit 10, which operates during the recognition phase of a speaker-dependent speech recognition system. Speech recognition unit 10 comprises template database 14, which in general will be template database 8 from training unit 6. The input to speech recognition unit 10 is digitized input speech $x(n)$, which is the speech
30 to be recognized. The input speech $x(n)$ is passed into parameter determination block 12, which performs the same parameter determination technique as parameter determination block 7 of training unit 6. Parameter determination block 12 produces a recognition template of N parameters $\{ t(n) \ n=1...N \}$ that models the characteristics of input speech $x(n)$.
35 Recognition template $t(n)$ is then passed to pattern comparison block 16 that performs a pattern comparison between template $t(n)$ and all the templates stored in template database 14. The distances between template $t(n)$ and each of the templates in template database 14 are forwarded to decision block 18, which selects from template database 14 the template that most closely

matches recognition template $t(n)$. The output of decision block 18 is the decision as to which word in the vocabulary was spoken.

Recognition accuracy is a measure of how well a recognition system correctly recognizes spoken words or phrases in the vocabulary. For example, a recognition accuracy of 95% indicates that the recognition unit correctly recognizes words in the vocabulary 95 times out of 100. In a traditional speech recognition system, the recognition accuracy is severely degraded in the presence of noise. The main reason for the loss of accuracy is that the training phase typically occurs in a quiet environment but the recognition typically occurs in a noisy environment. For example, a handsfree carkit speech recognition system is usually trained while the car is sitting in a garage or parked in the driveway, so the engine and air conditioning are not running and the windows are usually rolled up. However, recognition is normally used while the car is moving, so the engine is running, there is road and wind noise present, the windows may be down, etc. As a result of the disparity in noise level between the training and recognition phases, the recognition template does not form a good match with any of the templates obtained during training. This increases the likelihood of a recognition error or failure.

FIG. 3 illustrates a speech recognition unit 20 which must perform speech recognition in the presence of noise. As shown in FIG. 3, summer 22 adds speech signal $x(n)$ with noise signal $w(n)$ to produce noise-corrupted speech signal $r(n)$. It should be understood that summer 22 is not a physical element of the system, but is an artifact of a noisy environment. The noise-corrupted speech signal $r(n)$ is input to parameter determination block 24, which produces noise-corrupted template $t1(n)$. Pattern comparison block 28 compares template $t1(n)$ with all the templates in template database 26, which was constructed in a quiet environment. Since noise-corrupted template $t1(n)$ does not exactly match any of the training templates, there is a high probability that the decision produced by decision block 30 may be a recognition error or failure.

SUMMARY OF THE INVENTION

The present invention is a system and method for the automatic recognition of spoken words or phrases in the presence of noise. Speaker-dependent speech recognition systems operate in two phases: a training phase and a recognition phase. In the training phase of a traditional speech

recognition system, a user is prompted to speak all the words or phrases in a specified vocabulary. The digitized speech samples for each word or phrase are processed to produce a template of parameters characterizing the spoken words. The output of the training phase is a library of such templates. In
5 the recognition phase, the user speaks a particular word or phrase to initiate a desired action. The spoken word or phrase is digitized and processed to produce a template, which is compared with all the templates produced during training. The closest match determines the action that will be performed. The main impairment limiting the accuracy of speech
10 recognition systems is the presence of noise. The addition of noise during recognition severely degrades recognition accuracy, because this noise was not present during training when the template database was produced. The invention recognizes the need to account for the particular noise conditions that are present at the time of recognition to improve recognition accuracy.

15 Instead of storing templates of parameters, the improved speech processing system and method stores the digitized speech samples for each spoken word or phrase in the training phase. The training phase output is therefore a digitized speech database. In the recognition phase, the noise characteristics in the audio environment are continually monitored. When
20 the user speaks a word or phrase to initiate recognition, a noise-compensated template database is constructed by adding a noise signal to each of the signals in the speech database and performing parameter determination on each of the speech plus noise signals. One embodiment of this added noise signal is an artificially-synthesized noise signal with
25 characteristics similar to that of the actual noise. An alternative embodiment is a recording of the time window of noise that occurred just before the user spoke the word or phrase to initiate recognition. Since the template database is constructed using the same type of noise that is present in the spoken word or phrase to be recognized, the speech recognition unit
30 can find a good match between templates, improving the recognition accuracy.

BRIEF DESCRIPTION OF THE DRAWINGS

35 The features, objects, and advantages of the present invention will become more apparent from the detailed description set forth below when taken in conjunction with the drawings in which like reference characters identify correspondingly throughout and wherein:

FIG. 1 is a block diagram of a training unit of a speech recognition system;

FIG. 2 is a block diagram of a speech recognition unit;

FIG. 3 is a block diagram of a speech recognition unit which performs
5 speech recognition on a speech input corrupted by noise;

FIG. 4 is a block diagram of an improved training unit of a speech recognition system; and

FIG. 5 is a block diagram of an exemplary improved speech recognition unit.

10

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

This invention provides a system and method for improving speech
15 recognition accuracy when noise is present. It takes advantage of the recent advances in computation power and memory integration and modifies the training and recognition phases to account for the presence of noise during recognition. The function of a speech recognition unit is to find the closest
20 match to a recognition template that is computed on noise-corrupted speech. Since the characteristics of the noise may vary with time and location, the invention recognizes that the best time to construct the template database is during the recognition phase.

FIG. 4 shows a block diagram of an improved training unit 40 of a speech recognition system. As opposed to the traditional training method
25 shown in FIG. 1, training unit 40 is modified to eliminate the parameter determination step. Instead of storing templates of parameters, digitized speech samples of the actual words and phrases are stored. Thus, training unit 40 receives as input speech samples $s(n)$, and stores digitized speech samples $s(n)$ in speech database 42. After training, speech database 42
30 contains M speech signals, where M is the number of words in the vocabulary. Whereas the previous system and method of performing parameter determination loses information about the speech characteristics by only storing speech parameters, this system and method may preserve all the speech information for use in the recognition phase.

35 FIG. 5 shows a block diagram of an improved speech recognition unit 50 for use in conjunction with training unit 40. The input to speech recognition unit 50 is noise corrupted speech signal $r(n)$. Noise-corrupted speech signal $r(n)$ is generated by summer 52 adding speech signal $x(n)$ with

noise signal $w(n)$. As before, summer 52 is not a physical element of the system, but is an artifact of a noisy environment.

Speech recognition unit 50 comprises speech database 60, which contains the digitized speech samples that were recorded during the training phase. Speech recognition unit 50 also comprises parameter determination block 54, through which noise-corrupted speech signal $r(n)$ is passed to produce noise-corrupted template $t1(n)$. As in a traditional voice recognition system, parameter determination block 54 may implement any of a number speech parameter determination techniques.

An exemplary parameter determination technique uses linear predictive coding (LPC) analysis techniques. LPC analysis techniques model the vocal tract as a digital filter. Using LPC analysis, LPC cepstral coefficients $c(m)$ may be computed to be the parameters for representing the speech signal. The coefficients $c(m)$ are computed using the following steps. First, the noise-corrupted speech signal $r(n)$ is windowed over a frame of speech samples by applying a window function $v(n)$:

$$y(n) = r(n)v(n) \quad 0 \leq n \leq N-1 \quad (1)$$

In the exemplary embodiment, the window function $v(n)$ is a hamming window and the frame size N is equal to 160. Next, the autocorrelation coefficients are computed on the windowed samples using the equation:

$$R(k) = \sum_{m=0}^{N-k} y(m)y(m+k) \quad k=1,2,...,P \quad (2)$$

25

In the exemplary embodiment, P , the number of autocorrelation coefficients to be computed, is equal to the order of the LPC predictor, which is 10. The LPC coefficients are then computed directly from the autocorrelation values using Durbin's recursion algorithm. The algorithm may be stated as follows:

$$1. \quad E^{(0)} = R(0), \quad i = 1 \quad (3)$$

$$35 \quad 2. \quad k_i = \left\{ R(i) - \sum_{j=1}^{i-1} \alpha_j^{(i-1)} R(i-j) \right\} / E^{(i-1)} \quad (4)$$

$$3. \quad \alpha_i^{(i)} = k_i \quad (5)$$

7

$$4. \quad \alpha_j^{(i)} = \alpha_j^{(i-1)} - k_i \alpha_{i-j}^{(i-1)} \quad 1 \leq j \leq i-1 \quad (6)$$

$$5. \quad E^{(i)} = (1 - k_i^2) E^{(i-1)} \quad (7)$$

5 6. If $i < P$ then goto (2) with $i = i + 1$. (8)

7. The final solution for the LPC coefficients is given as

$$a_j = \alpha_j^{(P)} \quad 1 \leq j \leq P \quad (9)$$

10

The LPC coefficients are then converted to LPC cepstral coefficients using the following equations:

$$c(0) = \ln(R(0)) \quad (10)$$

$$15 \quad c(m) = a_m + \sum_{k=1}^{m-1} \left(\frac{k}{m} \right) c_k a_{m-k} \quad 1 \leq m \leq P \quad (11)$$

$$c(m) = \sum_{k=1}^{m-1} \left(\frac{k}{m} \right) c_k a_{m-k} \quad m > P \quad (12)$$

It should be understood that other techniques may be used for parameter determination instead of the LPC cepstral coefficients.

20 In addition, the signal $r(n)$ is passed to speech detection block 56 which determines the presence or absence of speech. Speech detection block 56 may determine the presence or absence of speech using any of a number of techniques. One such method is disclosed in the aforementioned U.S. Patent No. 5,414,796, entitled "VARIABLE RATE VOCODER." This
25 technique analyzes the level of speech activity to make the determination regarding the presence or absence of speech. The level of speech activity is based on the energy of the signal in comparison with the background noise energy estimate. First, the energy $E(n)$ is computed for each frame, which in a preferred embodiment is composed of 160 samples. The background noise
30 energy estimate $B(n)$ may then be calculated using the equation:

$$B(n) = \min [E(n), 5059644, \max (1.00547 * B(n-1), B(n-1) + 1)]. \quad (13)$$

If $B(n) < 160000$, three thresholds are computed using $B(n)$ as follows:

35

$$T1(B(n)) = -(5.544613 \times 10^{-6}) * B^2(n) + 4.047152 * B(n) + 362 \quad (14)$$

$$T2(B(n)) = -(1.529733 \times 10^{-5}) * B^2(n) + 8.750045 * B(n) + 1136 \quad (15)$$

$$T3(B(n)) = -(3.957050 \times 10^{-5}) * B^2(n) + 18.89962 * B(n) + 3347 \quad (16)$$

If $B(n) > 160000$, the three thresholds are computed as:

$$T1(B(n)) = -(9.043945 \times 10^{-8}) * B^2(n) + 3.535748 * B(n) - 62071 \quad (17)$$

$$T2(B(n)) = -(1.986007 \times 10^{-7}) * B^2(n) + 4.941658 * B(n) + 223951 \quad (18)$$

$$T3(B(n)) = -(4.838477 \times 10^{-7}) * B^2(n) + 8.630020 * B(n) + 645864 \quad (19)$$

This speech detection method indicates the presence of speech when energy $E(n)$ is greater than threshold $T2(B(n))$, and indicates the absence of speech when energy $E(n)$ is less than threshold $T2(B(n))$. In an alternative embodiment, this method can be extended to compute background noise energy estimates and thresholds in two or more frequency bands. Additionally, it should be understood that the values provided in Equations (13)-(19) are experimentally determined, and may be modified depending on the circumstances.

When speech detection block 56 determines that speech is absent, it sends a control signal that enables noise analysis, modeling, and synthesis block 58. It should be noted that in the absence of speech, the received signal $r(n)$ is the same as the noise signal $w(n)$.

When noise analysis, modeling, and synthesis block 58 is enabled, it analyzes the characteristics of noise signal $r(n)$, models it, and synthesizes a noise signal $w1(n)$ that has similar characteristics to the actual noise $w(n)$. An exemplary embodiment for performing noise analysis, modeling, and synthesis is disclosed in U.S. Pat. No. 5,646,991, entitled "NOISE REPLACEMENT SYSTEM AND METHOD IN AN ECHO CANCELLER," which is assigned to the assignee of the present invention and incorporated by reference herein. This method performs noise analysis by passing the noise signal $r(n)$ through a prediction error filter given by:

$$A(z) = 1 - \sum_{i=1}^P a_i z^{-i} \quad (20)$$

where P , the order of the predictor, is 5 in the exemplary embodiment. The LPC coefficients a_i , are computed as explained earlier using equations (1) through (9). Once the LPC coefficients are obtained, synthesized noise samples can be generated with the same spectral characteristics by passing white noise through the noise synthesis filter given by:

$$\frac{1}{A(z)} = \frac{1}{1 - \sum_{i=1}^P a_i z^{-i}} \quad (21)$$

which is just the inverse of the filter used for noise analysis. After applying a scaling factor to each of the synthesized noise samples to make the synthesized noise energy equal to the actual noise energy, the output is the synthesized noise $w1(n)$.

The synthesized noise $w1(n)$ is added to each set of digitized speech samples in speech database 60 by summer 62 to produce sets of synthesized noise corrupted speech samples. Then, each set of synthesized noise corrupted speech samples is passed through parameter determination block 64, which generates a set of parameters for each set of synthesized noise corrupted speech samples using the same parameter determination technique as that used in parameter determination block 54. Parameter determination block 54 produces a template of parameters for each set of speech samples, and the templates are stored in noise-compensated template database 66. Noise-compensated template database 66 is a set of templates that is constructed as if traditional training had taken place in the same type of noise that is present during recognition. Note that there are many possible methods for producing estimated noise $w1(n)$ in addition to the method disclosed in U.S. Pat. No. 5,646,991. An alternative embodiment is to simply record a time window of the actual noise present when the user is silent and use this noise signal as the estimated noise $w1(n)$. The time window of noise recorded right before the word or phrase to be recognized is spoken is an exemplary embodiment of this method. Still another method is to average various windows of noise obtained over a specified period.

Referring still to FIG. 5, pattern comparison block 68 compares the noise corrupted template $t1(n)$ with all the templates in noise compensated template database 66. Since the noise effects are included in the templates of noise compensated template database 66, decision block 70 is able to find a good match for $t1(n)$. By accounting for the effects of noise in this manner, the accuracy of the speech recognition system is improved.

The previous description of the preferred embodiments is provided to enable any person skilled in the art to make or use the present invention. The various modifications to these embodiments will be readily apparent to those skilled in the art, and the generic principles defined herein may be applied to other embodiments without the use of the inventive faculty.

Thus, the present invention is not intended to be limited to the embodiments shown herein but is to be accorded the widest scope consistent with the principles and novel features disclosed herein.

5 WE CLAIM:

CLAIMS

1. A speech recognition system, comprising:
 - 2 a training unit for receiving signals of words or phrases to be trained, generating digitized samples for each said words or phrases, and storing said
 - 4 digitized samples in a speech database; and
 - 6 a speech recognition unit for receiving a noise corrupted input signal to be recognized, generating a noise compensated template database by applying the effects of noise to said digitized samples of said speech database, and providing a speech recognition outcome for said noise corrupted input signal based on said noise compensated template database.
2. The speech recognition system of claim 1 wherein said speech
 - 2 recognition unit comprises:
 - 4 a first parameter determination unit for receiving said noise corrupted input signal and generating a template of parameters representative of said input signal in accordance with a predetermined
 - 6 parameter determination technique;
 - 8 a second parameter determination unit for receiving said speech database with the effects of noise applied to said digitized samples, and generating said noise compensated template database in accordance with
 - 10 said predetermined parameter determination technique; and
 - 12 a pattern comparison unit for comparing said template of parameters representative of said input signal with the templates of said noise compensated template database to determine the best match and thereby
 - 14 identify said speech recognition outcome.
3. The speech recognition system of claim 1 wherein said speech
 - 2 recognition unit comprises:
 - 4 a speech detection unit for receiving said noise corrupted input signal and determining whether speech is present in said input signal, wherein said input signal is designated a noise signal when speech is determined not
 - 6 to be present in said input signal; and
 - 8 a noise unit activated upon determining that speech is not present in said input signal, said noise unit for analyzing said noise signal and synthesizing a synthesized noise signal having characteristics of said noise
 - 10 signal, said synthesized noise signal for applying the effects of noise to said digitized samples of said speech database.

4. The speech recognition system of claim 2 wherein said speech
2 recognition unit further comprises:

a speech detection unit for receiving said noise corrupted input signal
4 and determining whether speech is present in said input signal, wherein
said input signal is designated a noise signal when speech is determined not
6 to be present in said input signal; and

a noise unit activated upon determining that speech is not present in
8 said input signal, said noise unit for analyzing said noise signal and
synthesizing a synthesized noise signal having characteristics of said noise
10 signal, said synthesized noise signal for applying the effects of noise to said
digitized samples of said speech database.

5. The speech recognition system of claim 2, wherein said
2 parameter determination technique is a linear predictive coding (LPC)
analysis technique.

6. The speech recognition system of claim 4, wherein said
2 parameter determination technique is a linear predictive coding (LPC)
analysis technique.

7. The speech recognition system of claim 3, wherein said speech
2 detection unit determines the presence of speech by analyzing the level of
speech activity in said input signal.

8. The speech recognition system of claim 4, wherein said speech
2 detection unit determines the presence of speech by analyzing the level of
speech activity in said input signal.

9. The speech recognition system of claim 3, wherein said noise
2 unit analyzes and synthesizes said synthesized noise signal using a linear
predictive coding (LPC) technique.

10. The speech recognition system of claim 3, wherein said
2 synthesized noise signal corresponds to a window of said noise signal
recorded right before said input signal to be recognized.

11. The speech recognition system of claim 3, wherein said
2 synthesized noise signal corresponds to an average of various windows of
said noise signal recorded over a predetermined period of time.

12. The speech recognition system of claim 4, wherein said noise
2 unit analyzes and synthesizes said synthesized noise signal using a linear
predictive coding (LPC) technique.

13. The speech recognition system of claim 4, wherein said
2 synthesized noise signal corresponds to a window of said noise signal
recorded right before said input signal to be recognized.

14. The speech recognition system of claim 4, wherein said
2 synthesized noise signal corresponds to an average of various windows of
said noise signal recorded over a predetermined period of time.

15. A training unit of a speech recognition system which accounts
2 for the effects of a noisy environment, comprising:
means for receiving signals of words or phrases to be trained;
4 means for generating digitized samples for each said words or
phrases; and
6 means for storing said digitized samples in a speech database.

16. A speech recognition unit of a speech recognition system for
2 recognizing an input signal, said speech recognition unit accounting for the
effects of a noisy environment, comprising:
4 means for storing digitized samples of words or phrases of a
vocabulary in a speech database;
6 means for applying the effects of noise to said digitized samples of
said vocabulary to generate noise corrupted digitized samples of said
8 vocabulary;
means for generating a noise compensated template database based on
10 said noise corrupted digitized samples; and
means for determining a speech recognition outcome for said input
12 signal based on said noise compensated template database.

17. The speech recognition unit of claim 16, further comprising:
2 first parameter determination means for receiving said input signal
and generating a template of parameters representative of said input signal

4 in accordance with a predetermined parameter determination technique;
and

6 second parameter determination means for receiving said noise
corrupted digitized samples of said vocabulary and generating the templates
8 of said noise compensated template database in accordance with said
predetermined parameter determination technique;

10 wherein said means for determining said speech recognition outcome
compares said template of parameters representative of said input signal
12 with the templates of said noise compensated template database to
determine the best match and thereby identify said speech recognition
14 outcome.

18. The speech recognition unit of claim 16 wherein said means
2 for applying the effects of noise comprises:

means for determining whether speech is present in said input signal,
4 wherein said input signal is designated a noise signal when speech is
determined not to be present in said input signal; and

6 means for analyzing said noise signal and synthesizing a synthesized
noise signal, said synthesized noise signal added to said digitized samples of
8 said vocabulary.

19. The speech recognition unit of claim 17 wherein said means
2 for applying the effects of noise comprises:

means for determining whether speech is present in said input signal,
4 wherein said input signal is designated a noise signal when speech is
determined not to be present in said input signal; and

6 means for analyzing said noise signal and synthesizing a synthesized
noise signal, said synthesized noise signal added to said digitized samples of
8 said vocabulary.

20. A method for speech recognition accounting for the effects of a
2 noisy environment, comprising the steps of:

generating digitized samples of each word or phrase trained, each said
4 word or phrase belonging to a vocabulary;

storing said digitized samples in a speech database;

6 receiving an input signal to be recognized;

applying the effects of noise to said digitized samples of said
8 vocabulary to generate noise corrupted digitized samples of said vocabulary;

generating a noise compensated template database based on said noise
10 corrupted digitized samples; and
providing a speech recognition outcome for said noise corrupted
12 input signal based on said noise compensated template database.

21. The method of speech recognition of claim 20, further
2 comprising the steps of:
generating a template of parameters representative of said input
4 signal in accordance with a predetermined parameter determination
technique; and
6 generating templates for said noise compensated template database in
accordance with said predetermined parameter determination technique;
8 wherein said step of providing a speech recognition outcome
compares said template of parameters representative of said input signal
10 with said templates of said noise compensated template database to
determine the best match and thereby identify said speech recognition
12 outcome.

22. The method of speech recognition of claim 20 wherein said
2 step of applying the effects of noise comprises the steps of:
determining whether speech is present in said input signal, wherein
4 said input signal is designated a noise signal when speech is determined not
to be present in said input signal; and
6 analyzing said noise signal and synthesizing a synthesized noise
signal, said synthesized noise signal added to said digitized samples of said
8 vocabulary to generate said noise corrupted digitized samples.

23. The method of speech recognition of claim 21 wherein said
2 step of applying the effects of noise comprises the steps of:
determining whether speech is present in said input signal, wherein
4 said input signal is designated a noise signal when speech is determined not
to be present in said input signal; and
6 analyzing said noise signal and synthesizing a synthesized noise
signal, said synthesized noise signal added to said digitized samples of said
8 vocabulary to generate said noise corrupted digitized samples.

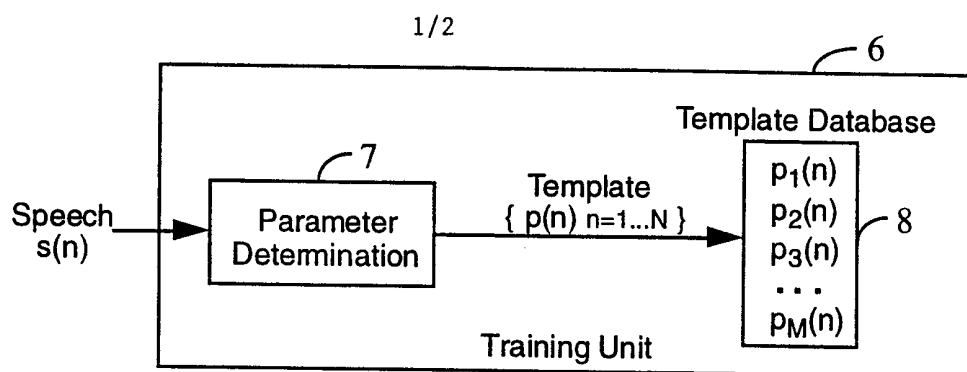


FIG. 1

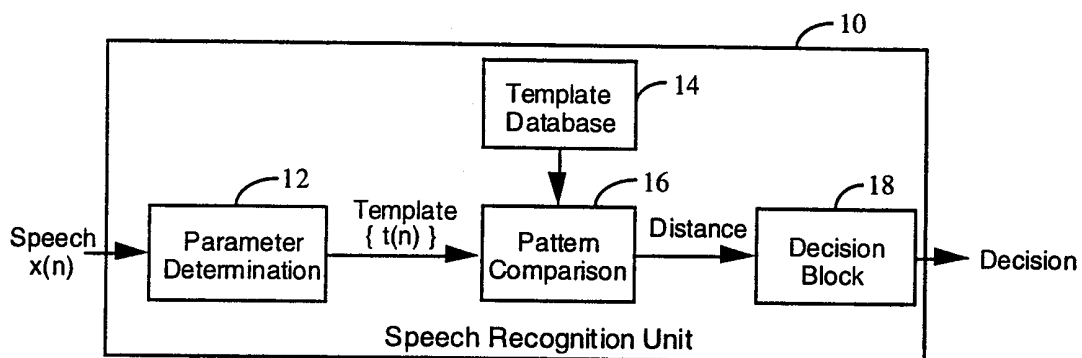


FIG. 2

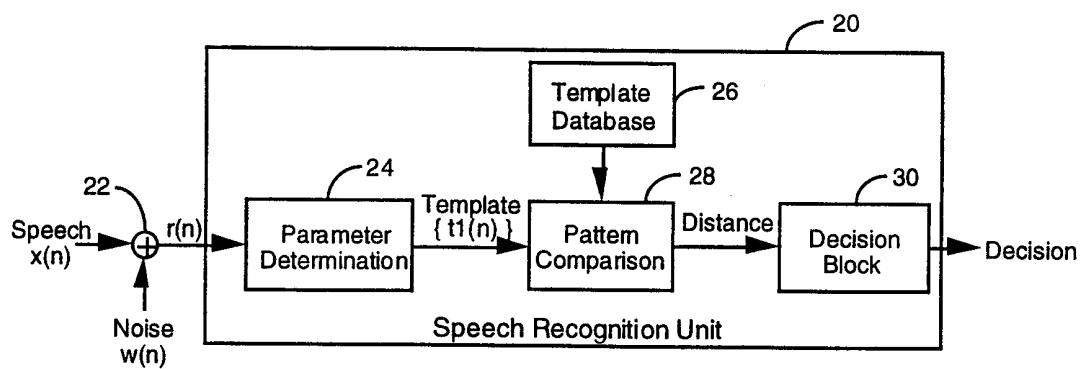


FIG. 3

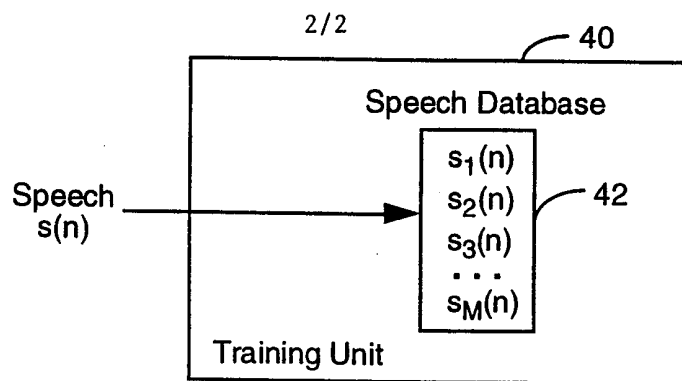


FIG. 4

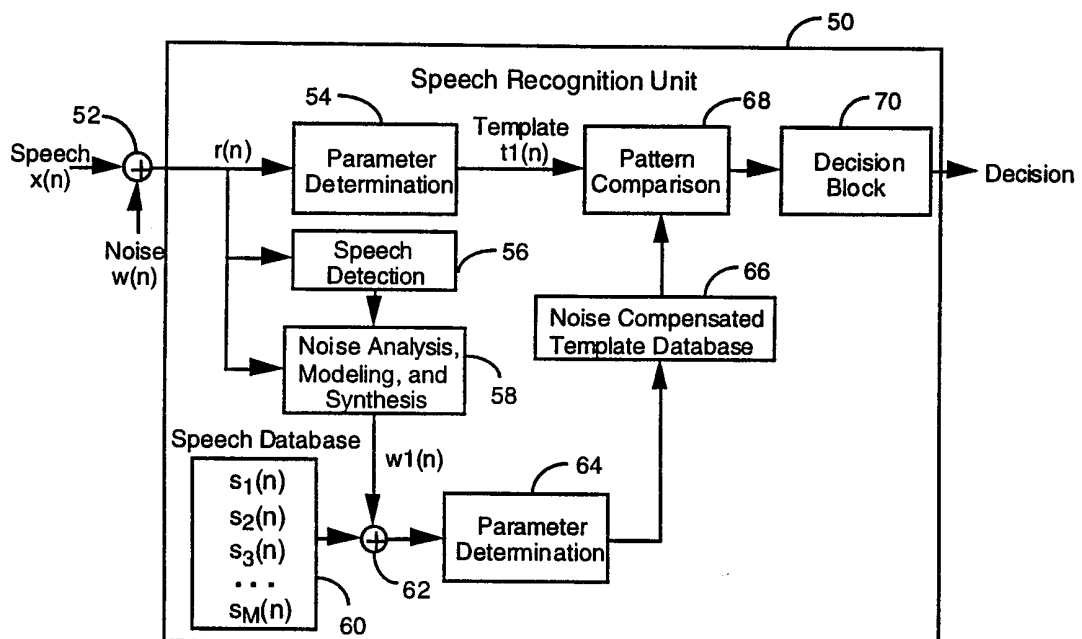


FIG. 5

INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 99/02280

A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 G10L5/06

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 G10L

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

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	GALES M J F ET AL: "ROBUST SPEECH RECOGNITION IN ADDITIVE AND CONVOLUTIONAL NOISE USING PARALLEL MODEL COMBINATION" COMPUTER SPEECH AND LANGUAGE, vol. 9, no. 4, 1 October 1995, pages 289-307, XP000640904 see page 292 ---	1,2,5, 15-17, 20,21
X	GONG Y: "Speech recognition in noisy environments: A survey" SPEECH COMMUNICATION, vol. 16, no. 3, 1 April 1995, page 261-291 XP004013163 Section 4.7 (Training data contamination), pages 280-281 --- -/--	1,2,5, 15-17, 20,21

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents:

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "&" document member of the same patent family

Date of the actual completion of the international search

4 May 1999

Date of mailing of the international search report

12/05/1999

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Ramos Sánchez, U

INTERNATIONAL SEARCH REPORT

Internat. Application No PCT/US 99/02280

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>US 4 933 973 A (PORTER JACK E) 12 June 1990 see figures 1A,4 see column 1, line 42 - line 51 see column 1, line 65 - column 2, line 7 see column 3, line 59 - column 4, line 5 see column 13, line 41 - line 62 see column 14, line 36 - column 15, line 22</p> <p style="text-align: center;">----</p>	1-23
A	<p>US 5 414 796 A (GARDNER WILLIAM R ET AL) 9 May 1995 cited in the application see abstract see column 6, line 49 - line 62 see column 14, line 61 - column 15, line 30</p> <p style="text-align: center;">----</p>	3-8,18, 19,22,23
A	<p>US 5 646 991 A (SIH GILBERT C) 8 July 1997 cited in the application</p> <p>see abstract see column 19, line 1 - line 14</p> <p style="text-align: center;">-----</p>	3-6,9, 12,18, 19,22,23

INTERNATIONAL SEARCH REPORT

Information on patent family members

Internati Application No

PCT/US 99/02280

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 4933973 A	12-06-1990	FR 2627887 A GB 2216320 A,B JP 1255000 A	01-09-1989 04-10-1989 11-10-1989
US 5414796 A	09-05-1995	US 5657420 A US 5778338 A AU 1482597 A AU 671952 B AU 2186592 A AU 693374 B AU 6089396 A BR 9206143 A CN 1071036 A CN 1159639 A CN 1167309 A EP 0588932 A FI 935597 A IL 102146 A IL 113986 A IL 113987 A IL 113988 A JP 6511320 T MX 9202808 A NO 934544 A WO 9222891 A ZA 9204082 A	12-08-1997 07-07-1998 15-05-1997 19-09-1996 12-01-1993 25-06-1998 03-10-1996 03-01-1995 14-04-1993 17-09-1997 10-12-1997 30-03-1994 13-12-1993 31-03-1996 14-11-1996 14-11-1996 14-11-1996 15-12-1994 01-12-1992 10-12-1993 23-12-1992 26-02-1993
US 5646991 A	08-07-1997	US 5559881 A US 5307405 A US 5687229 A AU 660243 B AU 5291893 A BG 61985 B BG 98780 A BR 9305647 A CN 1085706 A,B CZ 9401271 A EP 0615674 A FI 942443 A HU 70720 A,B IL 107100 A JP 7505037 T MX 9305889 A NO 941904 A PL 173748 B SK 60694 A WO 9408418 A ZA 9306322 A	24-09-1996 26-04-1994 11-11-1997 15-06-1995 26-04-1994 30-11-1998 28-08-1995 24-12-1996 20-04-1994 15-12-1994 21-09-1994 25-05-1994 30-10-1995 14-11-1996 01-06-1995 31-05-1994 22-07-1994 30-04-1998 09-11-1994 14-04-1996 05-05-1994