

CONTINUOUS MANDARIN CHINESE SPEECH RECOGNITION SYSTEM
HAVING AN INTEGRATED TONE CLASSIFIER

CROSS-REFERENCE TO RELATED APPLICATIONS

The present invention relates to pending U.S. Patent Application Serial No.:
, filed _____, 1994, invented by Hsiao-Wuen Hon and Bao-Sheng Yuan, entitled "A
System And Method For Recognizing A Tonal Language," which is incorporated
herein by reference. The present invention also relates to pending U.S. Patent
Application Serial No.: _____, filed _____, 1994, invented by Hsiao-Wuen Hon,
entitled "A System And Method For Determining The Tone Of A Syllable Of Mandarin
Chinese Speech," which is incorporated herein by reference.

BACKGROUND OF THE INVENTION

1. Field of the Invention.

The present invention relates generally to speech recognition systems. In
particular, the present invention relates to a system and method for recognizing
continuous Mandarin Chinese speech. The present invention further relates to a
system and method for recognizing continuous speech of a tonal language using an
integrated tone classifier.

2. Description of the Background Art.

Speech recognition systems for Mandarin Chinese and other tonal languages
encounter unique problems not encountered by speech recognition systems for non-
tonal languages such as Romance or Germanic languages. Mandarin Chinese is a tonal
syllabic language where each syllable is assigned a tone. In Mandarin Chinese, there
are 4 lexical tones (high and level, rising, falling-rising, and falling) and 1 neutral tone.
Tone is characterized by the fundamental frequency contour, or pitch contour, of the
audio signal. The pitch is equivalent to the fundamental frequency, and the pitch
contour is equivalent to the fundamental frequency contour. Exemplary wave forms
for electrical signals representing the tones of Mandarin Chinese are shown in Figures

1 1A, 1B, 1C, 1D, and 1E. The tone and syllable together define the meaning of the
2 syllable. Syllables with the same phonetic structure, but different tones, usually have
3 significantly different meanings. Thus, to recognize accurately an audio signal of
4 Mandarin Chinese speech, a speech recognition system must recognize both the
5 syllable and the tone of the syllable.

6 There are many prior art systems, similar to systems for non-tonal languages,
7 that effectively analyze and identify isolated syllables of Mandarin Chinese speech.
8 These systems have been quite successful in accurately recognizing isolated syllables.
9 Such prior art systems generally first identify the syllable and second perform a tonal
10 analysis. The systems then combine the results of the two steps to recognize the input.

11 Prior art systems that recognize continuous Mandarin Chinese speech have not
12 been nearly as successful as systems that recognize isolated syllables. Continuous
13 speech recognition systems must recognize both the syllable and tone of each of a
14 plurality of syllables strung together in a continuous input.

15 Existing continuous speech recognition systems first divide or segment an input into a
16 sequence of fixed segments with hypothetical time alignment. The step of segmenting
17 the input is particularly critical since errors in segmentation will propagate through
18 and affect the recognition of both syllable and tone. There are, however, no
19 segmentation techniques that correctly segment continuous Mandarin Chinese speech
20 such that this approach yields enough accuracy to be satisfactory.

21 Once these prior art systems have segmented the input, they generally use an
22 isolated syllable recognizer and separate tone recognizer to identify each tonal syllable
23 based on the hypothetical segments. This analysis is obviously dependent upon the
24 segmentation step. These systems have an additional problem in that they use short-
25 term tonal analysis which does not provide sufficient frequency resolution to identify
26 correctly the behavior of the pitch (the fundamental frequency) contour. Moreover, the
27 tone of a syllable may move through 3 octaves or more from one syllable to the next.
28 In order to overcome the deficiencies of short-term tonal analysis and the difficult
29 behavior of the tone, long-term tonal analysis is needed to model accurately the pitch

1 contour. Long-term tonal analysis, however, is very sensitive to segmentation error.
2 Furthermore, long-term tonal analysis is also is very time consuming. Time
3 consumption is particularly important when a speech recognition system is being used
4 for real-time speech applications.

5 In an attempt to reduce the effects of segmentation error on the ultimate
6 recognition results, prior art continuous speech recognition systems provide multiple
7 possible identifications of an input. Such systems determine multiple candidates for an
8 input utterance and generate output signals of the N-best candidates or recognitions.
9 A recognition is conventionally referred to as a theory. These systems often generate
10 an initial confidence score with each possible recognition. Each initial confidence score
11 is an indication of how accurately the theory matches the input. Generally, the
12 recognition with the highest initial confidence score is accepted as the correct
13 recognition. These prior art speech recognition systems have not utilized long-term
14 tonal analysis because of the computational expense. Long-term tonal analysis is even
15 more time consuming when it must be performed for each of the N-best theories.

16 Continuous speech recognition systems that produce multiple possible
17 recognitions, still encounter the problem of incorporating long-term tonal analysis into
18 an N-best recognizer. Thus, there continues to be a need for performing long-term
19 tonal analysis, with minimal degradation due to segmentation error, for continuous
20 speech of a tonal language.

21

22 SUMMARY OF THE INVENTION

23 The present invention overcomes the deficiencies and limitations of the prior art
24 with a system and method for recognizing continuous Mandarin Chinese speech. The
25 system comprises a microphone, an A/D converter, a syllable recognition system, an
26 integrated tone classifier, and a confidence score augmentor. The system receives
27 audio signals through the microphone which converts the audio signal into an analog,
28 electrical signal. The A/D converter digitizes the electrical signal.

1 The syllable recognition system analyzes an input of continuous Mandarin
2 Chinese speech and generates N-best theories. Each theory is divided into one or more
3 syllables. The syllable recognition system determines the phonetic structure of each
4 syllable of each segmentation pattern and might perform short-term tonal analysis to
5 determine the tone of each syllable of each theory. A segmentation pattern is a
6 possible division of the input into discrete syllables. The syllable recognition system
7 combines the identified syllables and tones of each segmentation pattern into a theory
8 and determines an initial confidence score for each theory. The initial confidence score
9 indicates the determination of the syllable recognition system of how well the phonetic
10 structure and tone of the theory match the phonetic structures and tones of the
11 language. The syllable recognition system ranks the theories by initial confidence score
12 and generates an output of the N-best theories with their initial confidence scores.

13 The integrated tone classifier comprises a pitch estimator and a long-term tone
14 analyzer. The pitch estimator receives the digitized input from the A/D converter and
15 estimates the pitch of the input. The integrated tone classifier estimates the pitch only
16 once; thus reducing the computation time required for long-term tonal analysis. The
17 long-term tone analyzer receives the estimated pitch from the pitch estimator and the
18 N-best theories from the syllable recognition system. The long-term tone analyzer
19 segments the estimated pitch according to the syllables of each of the theories and
20 determines how well the estimated pitch, as segmented, matches the tones of the
21 language. The long-term tone analyzer generates a long-term tonal confidence signal
22 for each theory. The long-term tonal confidence signal indicates how well the syllables
23 of the theory match the tones of the language. The integrated tone classifier outputs
24 the long-term tonal confidence signals to the confidence score augmentor.

25 The confidence score augmentor receives the N-best theories and initial
26 confidence scores and the long-term tonal confidence signals from integrated tone
27 classifier. The confidence score augmentor modifies the initial confidence score of each
28 theory according to its corresponding long-term tonal confidence signal. The
29 confidence score augmentor re-ranks the theories according to the augmented

1 confidence scores and generates an output of the re-ranked theories with augmented
2 confidence scores.

3 The present invention also includes a method for recognizing an input of
4 continuous Mandarin Chinese speech. The method begins by generating N-best
5 theories using syllable recognition, also referred to as phonetic identification, and
6 possibly includes short-term tonal analysis techniques. The method estimates the pitch
7 of the input and then segments the estimated pitch into units according to the syllable
8 segmentation of a theory. Each unit is then analyzed using long-term tonal analysis
9 techniques and a long-term tonal confidence signal is generated for each theory. The
10 method finally modifies the initial confidence scores, generated by the phonetic
11 identification and short-term tonal analysis, according to their corresponding long-
12 term tonal confidence signal and re-ranks the theories.

13

14

BRIEF DESCRIPTION OF THE DRAWINGS

15 Figures 1A, 1B, 1C, 1D, and 1E are graphical representations of electrical signals
16 having the behaviors of the 4 lexical tones and one neutral tone of Mandarin Chinese
17 speech;

18 Figure 2 is a block diagram of a first embodiment of the system of the present
19 invention for recognizing continuous Mandarin Chinese;

20 Figure 3 is a block diagram of a second embodiment of the system of the present
21 invention for recognizing continuous Mandarin Chinese speech;

22 Figure 4 is a block diagram of a preferred embodiment of the speech recognition
23 unit of the second embodiment of the present invention; and

24 Figure 5 is a flowchart of a preferred method of the present invention for
25 recognizing continuous Mandarin Chinese speech.

26

27

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

28 Referring now to Figures 1A, 1B, 1C, 1D, and 1E, graphical representations of
29 the fundamental frequency behavior of electrical signals that represent the 4 lexical

tones and 1 neutral tone of Mandarin Chinese speech are shown. The signals shown in Figures 1A, 1B, 1C, 1D, and 1E originate as audio signals generated by a human speaker. A speech recognition system 76 receives the audio signals through a microphone 40. The microphone 40 converts the audio signal to an electrical signal. Each syllable of Mandarin Chinese speech has a tone. The tone is the behavior of the fundamental frequency contour, or pitch, of the audio signal of the syllable. The tone is preserved when the audio signal is converted into an electrical signal. Thus, the tone of a syllable may refer to the behavior of the fundamental frequency contour of the audio signal or the behavior of the fundamental frequency of the electrical signal into which the audio signal is converted.

Figures 1A, 1B, 1C, 1D, and 1E show the tones of Mandarin Chinese speech as the behavior of the fundamental frequency plotted as a function of time. Figure 1A shows a first tone signal 90 having a high and level tone. Figure 1B shows a second tone signal 92 having a rising tone. A third tone signal 94, shown in Figure 1C, has a falling-rising behavior. Figure 1D shows a fourth tone signal 96 having a behavior of the last lexical tone of Mandarin Chinese speech, the falling tone. Finally, Figure 1E shows a fifth tone signal 98 having the neutral tone.

Referring now to Figure 2, a block diagram of a first embodiment of a system 76 for recognizing continuous Mandarin Chinese speech is shown. The system 76 comprises a microphone 40, an analog-to-digital converter 42, a syllable recognition system 50, an integrated tone classifier 58, and a confidence score augmentor 56. The system 76 receives an input of an audio signal of continuous Mandarin Chinese speech through the microphone 40. The microphone 40 is a conventional microphone for converting audio signals into electrical signals. The microphone 40 has an input for receiving audio signals and an output for transmitting analog, electrical signals. The microphone 40 transforms the audio signal into an analog electrical signal and transmits the analog electrical signal on a line 60 that is coupled to the output of the microphone 40. The line 60 is also coupled to an input of the analog-to-digital converter 42. The analog-to-digital converter ("A/D converter") 42 is a conventional

1 A/D converter for digitizing analog, electrical signals. The A/D converter 42 has a
2 single input, coupled to the line 60, and a single output, coupled to a line 62. The A/D
3 converter 42 converts the input analog, electrical signal into a digital, electrical signal
4 and outputs the digital, electrical signal on the line 62.

5 The line 62 is coupled to an input of the syllable recognition system 50 and to a
6 first input of the integrated tone classifier 58. The syllable recognition system 50
7 receives the digitized output signal from the A/D converter 42 and generates N-best
8 theories with initial confidence scores. The syllable recognition system 50 asserts the
9 theories and confidence scores on a line 68, which is coupled to an output of the
10 syllable recognition system 50. Each output comprises at least one theory, which is a
11 possible recognition (a sequence of syllables with time segmentation) of the input
12 signal, and an initial confidence score, which indicates the confidence of the syllable
13 recognition system 50 in correctly recognizing the input signal. There are many ways
14 that the initial confidence score may indicate the confidence of the syllable recognition
15 system 50 that it accurately recognized the input signal. For example, the confidence
16 scores may take on values in a range such as from 0 to 100. In this case, a value of 100
17 represents the maximum confidence of the syllable recognition system 50, and initial
18 confidence scores with greater value represent greater confidence. Alternatively, the
19 initial confidence scores may take on values between 0 and 1 with 0 representing the
20 maximum confidence of the syllable recognition system 50. In this case, the lower the
21 initial confidence score the greater the confidence level. There are many different
22 techniques for generating initial confidence scores; those skilled in the art will
23 recognize these techniques. The syllable recognition system 50 preferably generates a
24 plurality of theories with accompanying initial confidence scores but outputs only the
25 N-best theories. N is a pre-defined value and is preferably set by the system developer
26 according to his/her application's requirements. The syllable recognition system 50 is
27 preferably an Apple Plaintalk Chinese Syllable Recognition System from Apple
28 Computer, Inc. of Cupertino, California. The syllable recognition system 50 may

alternately be a syllable based voice typewriter from Star Incorporated of the People's Republic of China.

Line 62, which provides the output of the A/D converter 42, is coupled to the first input of the integrated tone classifier 58. Line 68 is coupled to a second input of the integrated tone classifier 58. Systems for Mandarin Chinese speech recognition, which include the integrated tone classifier 58 of the present invention, overcome the deficiencies and limitations of the prior art speech recognition systems. In particular, the integrated tone classifier 58 quickly and efficiently applies long-term tonal analysis to the output of the syllable recognition system 50. Long-term tonal analysis allows the system 76 to determine accurately the tone of the input and thus, greatly improves the accuracy of the recognition of an input audio signal.

The integrated tone classifier 58 preferably comprises a pitch estimator 52 and a long-term tone analyzer 54. The pitch estimator 52 estimates the pitch, or fundamental frequency, of the entire digitized input signal. The pitch estimator 52 has a single input that forms the first input of the integrated tone classifier 58 and that is coupled to line 62 to receive the digitized input signal. The pitch estimator 52 determines the pitch of the entire input signal and provides a pitch signal to the long-term tone analyzer 54. The pitch estimator 52 uses fourier analysis to determine the behavior of the fundamental frequency of the input signal. The pitch estimator 52 may alternately use a low pass filter to determine the pitch of the input signal. There are many other techniques that may be used to determine the pitch of the input signal. Those skilled in the art will recognize these techniques. The pitch estimator 52 preferably determines the pitch of the input signal only once. In this way, the present invention eliminates the repeated determination of the pitch for the N-best theories of the syllable recognition system 50 which is computationally expensive. The present invention advantageously performs pitch estimation first for the entire input utterance and then divides the pitch signal in different ways according to the segmentation of each of the N-best theories. Thus, the present invention overcomes the deficiency and limitation of undue delay of the prior art systems that attempted to apply long-term tonal analysis.

1 The output of the pitch estimator 52 is coupled to a first input of the long-term
2 tone analyzer 54 by a line 70 so that the long-term tone analyzer 54 receives the
3 estimated pitch signal. A second input of the long-term tone analyzer 54 forms the
4 second input of the integrated tone classifier 58 and is coupled to the line 68 to receive
5 the N-best theories, with their initial confidence scores, from the syllable recognition
6 system 50. The long-term tone analyzer 54 includes a memory device for storing the
7 estimated pitch signal from the pitch estimator 52. The long-term tone analyzer 54 is
8 preferably the system described in co-pending U.S. patent application entitled "A
9 System And Method For Determining The Tone Of A Syllable Of Mandarin Chinese
10 Speech" which is incorporated herein by reference. The long-term tone analyzer 54
11 may alternately be a system as described by Chih-Heng Lin, Lin-Shan Lee, and Pei-Yih
12 Ting in "A New Framework For Recognition Of Mandarin Syllables With Tones Using
13 Sub-Syllabic Units. Published in the Proceedings of the Institute of Electrical and
14 Electronic Engineers, 1993. The long-term tone analyzer 54 segments the estimated
15 pitch signal into units according to each of the N-best theories from the syllable
16 recognition system 50. Each unit is equal in duration to a corresponding syllable of a
17 theory. For example, for the highest ranked theory, the long-term tone analyzer 54
18 divides the pitch signal into the same number of units as the number of syllables
19 contained in the theory. Each unit has a duration equal to its corresponding syllable.
20 Thus, the present invention reduces the effect of segmentation error by using the
21 syllable recognition system 50 as a sophisticated segmenter and allows for different
22 segmentations for each of the N-best theories. The output of the syllable recognition
23 system 50 is the N-best theories that the syllable recognition system 50 was able to
24 determine. By segmenting the pitch signals according to the syllables of the N-best
25 output, the long-term tone analyzer 54 uses segmentation patterns that the syllable
26 recognition system 50 has already determined to be the N-best matches. The
27 integrated tone classifier 58 performs long-term tonal analysis on the input audio
28 signal that is optimally segmented. This greatly improves the accuracy of the
29 recognition.

1 The long-term tone analyzer 54 then compares each unit to the 4 lexical tones
2 and one neutral tone of Mandarin Chinese speech and generates a long-term
3 confidence signal that indicates how well the units match the tones. The long-term
4 tone analyzer 54 preferably uses the system described in co-pending U.S. patent
5 application entitled "A System And Method For Determining The Tone Of A Syllable
6 Of Mandarin Chinese Speech," which is incorporated herein by reference. The long-
7 term confidence signal is normalized so that the signals of theories having different
8 numbers of syllables can be compared. Other tonal languages have different numbers
9 of tones; for these languages, the long-term tone analyzer 54 would compare the units
10 to all tones of the tonal language. The long-term tone analyzer 54 then outputs the
11 long-term tonal confidence signals for each of the N-best theories on a line 72 that is
12 coupled to its output. The output of the long-term tone analyzer 54 forms the output
13 of the integrated tone classifier 58.

14 The line 72 is coupled to a first input of a confidence score augmentor 56. The
15 line 68, which is coupled to the output of the syllable recognition system 50 is also
16 coupled to a second input of the confidence score augmentor 56. The confidence score
17 augmentor 56 modifies the initial confidence scores received from the syllable
18 recognition system 50 according to the long-term tonal confidence signals generated by
19 the integrated tone classifier 58. For example, the confidence score augmentor 56 may
20 adjust the initial confidence scores by adding the corresponding long-term tonal
21 confidence signal to each of the initial confidence scores. More specifically, the
22 augmented confidence score output of the confidence score augmentor 56 may be a
23 signal that is the sum of 50% of the initial confidence score and 50% of the long-term
24 confidence signal. The percentage weighting for the initial confidence score and the
25 long-term confidence signal may be adjusted by the system developer as desired to
26 account for the weighting that the integrated tone classifier 58 will provide in the
27 overall recognition. There are many different techniques for combining the initial
28 confidence scores and the long-term tonal confidence signals; those skilled in the art
29 will recognize these techniques. The confidence score augmentor 56 then re-ranks the

1 N-best theories according to their augmented confidence scores and outputs the results
2 on a line 74. Line 74 may be coupled to a look up table and from there to a memory, as
3 in a dictation machine, or to a display device, or to both.

4 Referring now to Figure 3, a block diagram of a second and preferred
5 embodiment of a system 10 for recognizing continuous Mandarin Chinese speech,
6 constructed in accordance with the present invention, is shown. The system 10
7 preferably comprises an input device 12, an output device 14, a central processing unit
8 (CPU) 16, a memory means 18, and a speech recognition unit 20. The input device 12,
9 output device 14, central processing unit (CPU) 16, and memory means 18, are coupled
10 in a von Neuman architecture via a bus 22 such as in a personal computer. The central
11 processing unit (CPU) 16 is preferably a microprocessor such as a Motorola 68040; the
12 output device 14 is preferably a video monitor; the memory means 18 comprises
13 random access memory ("RAM") and read-only memory ("ROM") for storing data and
14 program instruction steps. The input device 12 comprises a keyboard, mouse type
15 controller, microphone, and A/D converter. In the preferred embodiment, the system
16 10 is a Macintosh Quadra 840AV computer system from Apple Computer, Inc. of
17 Cupertino, California. Those skilled in the art will realize that the system 10 could also
18 be implement on an IBM personal computer or other computer systems.

19 In the preferred embodiment, the speech recognition unit 20 is a memory,
20 coupled to the bus 22, that stores a set of program instruction steps that when executed
21 by the CPU 16 recognizes continuous Mandarin Chinese speech. The system 10 of the
22 present invention receives audio signal inputs of continuous Mandarin Chinese speech
23 through the microphone of the input device 12. The speech recognition unit 20
24 analyzes the input and produces N-best recognitions of the input. The speech
25 recognition unit 20 preferably receives an input, analyzes the input, and generates a
26 plurality of output signals of possible recognitions of the input signal. Each output
27 signal includes a signal of the recognized input, referred to as a theory, and an initial
28 confidence score. The initial confidence score indicates the confidence of the speech
29 recognition unit 20 that the associated theory is a correct recognition of the input and is

1 adjusted by a confidence score augmentor 24, shown in Figure 4, according to a signal
2 from an integrated tone classifier 32. Preferably, the speech recognition unit 20
3 generates one or more theories and ranks the theories by initial confidence score.

4 Referring now to Figure 4, a block diagram of the speech recognition unit 20 is
5 shown. The speech recognition unit 20 comprises a confidence score augmentor 24, a
6 syllable recognition system 30, and the integrated tone classifier 32. The syllable
7 recognition system 30 receives the digitized input signal and generates N-best theories
8 with accompanying initial confidence scores. The syllable recognition system 30 is
9 preferably the Apple Plaintalk Chinese Syllable Recognition System from Apple
10 Computer, Inc. of Cupertino, California. As was discussed with reference to Figure 2,
11 the syllable recognition system 30 determines the phonetic structure of each portion of
12 each segmentation pattern and might perform short-term tonal analysis. The syllable
13 recognition system 30 generates an initial confidence score for each theory and ranks
14 the theories by initial confidence score. The syllable recognition system 30 then
15 generates a signal to transmit the N-best theories, with accompanying initial
16 confidence scores, on the bus 22 to the integrated tone classifier 32 and to the
17 confidence score augmentor 24.

18 The integrated tone classifier 32 is coupled to the bus 22 to receive the digitized
19 input signal from the A/D converter and the N-best theories from the syllable
20 recognition system 30. The integrated tone analyzer 32 comprises a pitch estimator 38
21 and a long-term tone analyzer 39. The pitch estimator 38 estimates the pitch of the
22 entire input signal and transmits the result to the long-term tone analyzer 39. The
23 pitch estimator 38 preferably estimates the pitch only once. By estimating the pitch
24 only once, the integrated tone classifier 32 greatly speeds the long-term analysis of the
25 N-best theories. The long-term tone analyzer 39 segments the estimated pitch
26 according to the syllables of the N-best theories generated by the syllable recognition
27 system 30, determines the tone of each syllable of each of the N-best theories, and
28 generates a long-term tonal confidence signal for each theory. The integrated tone

1 classifier 32 has an output coupled to the bus 22 through which it asserts the long-term
2 tonal confidence signal in the bus 22 to the confidence score augmentor 24.

3 The confidence score augmentor 24 has a first input and a second input, both
4 coupled to the bus 22 to receive the long-term tonal confidence signals from the
5 integrated tone classifier 32 and the N-best theories, with associated initial confidence
6 scores, from the syllable recognizer 30, respectively. The confidence score augmentor
7 24 modifies each initial confidence score according to the corresponding long-term
8 tonal confidence signal, re-ranks the theories according to their augmented confidence
9 scores, and generates a signal to output the N-best theories with augmented confidence
10 scores. The confidence score augmentor 24 is coupled to the bus 22 to output the N-
11 best theories with augmented confidence scores.

12 Referring now to Figure 5, a flow chart of a preferred method for analyzing an
13 input of continuous Mandarin Chinese speech using the present invention is shown.
14 The method begins in step 100 where the system 10 receives an input of continuous
15 Mandarin Chinese speech. In step 102, the microphone of the input device 12 converts
16 the audio signal into an analog, electrical signal, and the A/D converter of the input
17 device 12 digitizes the input signal. In step 104, the syllable recognition system 30
18 processes the input and generates multiple theories with corresponding initial
19 confidence scores. The syllable recognition system 30 outputs the N-best theories
20 according to the initial confidence scores. N is a number selected by the system
21 developer according to his/her application's requirements. If the syllable recognition
22 system 30 can not generate more than N theories for an input, it outputs all the
23 theories, ranked by initial confidence score, that it can generate.

24 In step 106, the pitch estimator 38 receives the digitized input signal from the
25 input device 12 and estimates the pitch of the input signal. The method preferably
26 estimates the pitch only once because estimating pitch is very computationally
27 expensive. The present invention overcomes the deficiencies of the prior art by using
28 long-term tonal analysis but avoids succumbing to computational overhead of
29 estimating the pitch multiple times. Those skilled in the art will realize that step 106

1 may be performed concurrently with step 104. The pitch estimator 38 uses fourier
2 analysis or a low pass filter to estimate the fundamental frequency of the input signal.
3 Those skilled in the art will realize that many other methods may be used to estimate
4 the pitch. In step 108, the long-term tone analyzer 39 receives the estimated pitch from
5 the pitch estimator 38 and stores the estimated pitch.

6 In step 110, the long-term tone analyzer 39 segments the estimated pitch into
7 units according to the syllable segmentation of each of the N-best theories. In this way,
8 the integrated tone classifier 32 uses the syllable recognition 30 as a sophisticated
9 segmenter to divide the input syllables into reliable syllable segments for long-term
10 tonal analysis. The syllable recognition system 30 provided the N-best theories that
11 already closely approximate the correct segmentation. By using the syllable
12 recognition system 30 for segmenting the estimated pitch into multiple syllable
13 segments, the present invention reduces the effects of segmentation error. The units
14 are then compared, in step 112, to the tones of the language by the long-term tone
15 analyzer 39. The long-term tone analyzer 54 preferably uses the method described in
16 co-pending U.S. patent application entitled "A System And Method For Determining
17 The Tone Of A Syllable Of Mandarin Chinese Speech," which is incorporated herein by
18 reference. The long-term tone analyzer 39 generates a long-term tonal confidence
19 signal for each theory in step 114. The long-term tonal confidence signal indicates how
20 well the units match the tones of the language. As with the other scores, the long-term
21 tonal confidence signal is normalized so that the long-term tonal confidence signals of
22 theories having different numbers of syllables may be compared.

23 The confidence score augmentor 24 then, in step 116, receives the theories, with
24 their initial confidence scores, from the syllable recognition system 30 and receives the
25 long-term tonal confidence signals from the integrated tone classifier 32. The
26 confidence score augmentor 24 modifies the initial confidence scores of each theory
27 according to the long-term tonal confidence signal that corresponds to the theory. The
28 confidence score augmentor 24 then re-ranks, in step 118, the theories according to the
29 augmented confidence scores. Finally, in step 120, the confidence score augmentor 24

1 generates an output signal for the speech recognition system 10 of the N-best theories
2 and their augmented confidence scores. The method then ends.

3 While the present invention has been described with reference to certain
4 preferred embodiments, those skilled in the art will recognize that various
5 modifications may be provided. It is not required that the initial confidence score and
6 tonal score indicate greater confidence or match by a more positive value. The initial
7 confidence score may be augmented in any manner, not only by addition of the long-
8 term tonal confidence signal to the initial confidence score. Furthermore, the present
9 invention may be used to recognize any tonal language, such as Cantonese, and is not
10 limited to Mandarin Chinese. The present invention may also be used with
11 handwriting recognition systems and other data recognition systems. These and other
12 variations upon and modifications to the preferred embodiments are provided for by
13 the present invention, which is limited only by the following claims.



1 WHAT IS CLAIMED IS:

2 1. An integrated tone classifier for performing long-term tonal analysis of
3 an input signal of continuous speech of a tonal language, the integrated tone classifier
4 comprising:

5 a pitch estimator, having an input coupled to receive the input signal and an
6 output, for estimating the pitch contour of the input signal; and
7 a long-term tone analyzer, having an input coupled to the output of the pitch
8 estimator and an output forming an output of the integrated tone
9 classifier, for segmenting an estimated pitch contour generated by the
10 pitch estimator into units and for performing long-term tonal analysis on
11 the units of the segmented, estimated pitch.

12
13 2. The system of claim 1, wherein the pitch estimator comprises:
14 a processor, having inputs and outputs coupled to the inputs and outputs of the
15 long-term tone analyzer, for executing program instruction steps; and
16 a first memory device, having inputs and outputs coupled to the inputs and
17 outputs of the processor, for storing program instruction steps for
18 execution by the processor to estimate the pitch contour of the input
19 signal.

20
21 3. The system of claim 2, wherein the long-term tone analyzer comprises a
22 second memory device, having inputs and outputs coupled to the inputs and outputs
23 of the processor and the inputs and outputs of the first memory device, for storing
24 program instruction steps for execution by the processor to segment an estimated pitch
25 contour generated by the pitch estimator into units and for performing long-term tonal
26 analysis on the units of the segmented, estimated pitch.

27
28 4. A system for recognizing an input signal of continuous speech of a tonal
29 language, the system comprising:



1 a syllable recognition system, having an input and an output, for determining
2 the phonetic structure of syllables, for performing short-term tonal
3 analysis of the input, and for generating N-best theories with initial
4 confidence scores, the input of the syllable recognition system coupled to
5 receive the input signal; and
6 an integrated tone classifier, having a first and second input and an output, the
7 first input coupled to receive the input signal and the second input
8 coupled to the output of the syllable recognition system to receive the N-
9 best theories with initial confidence scores, for performing long-term
10 tonal analysis to determine the tone of syllables of the theories and for
11 generating a long-term tonal confidence signal for each theory.

12
13 5. The system of claim 1, wherein the pitch estimator uses fourier analysis to
14 determine the pitch contour of the input signal.

15
16 6. The system of claim 1, wherein the pitch estimator uses a low pass filter
17 to determine the pitch contour of the input signal.

18
19 7. The system of claim 4, wherein the integrated tone classifier comprises:
20 a pitch estimator, having an input coupled to receive the input signal and an
21 output, for estimating the pitch contour of the input signal; and
22 a long-term tone analyzer, having a first input coupled to an output of the pitch
23 estimator and a second input coupled to an output of the syllable
24 recognition system and an output forming the output of the integrated
25 tone classifier, for segmenting an estimated pitch contour generated by
26 the pitch estimator into units according to an output of the syllable
27 recognition system and for performing long-term tonal analysis on the
28 units of the segmented, estimated pitch.

29



1 8. The system of claim 7, wherein the long-term tone analyzer generates a
2 long-term tonal confidence signal indicating how well the units of a segmented,
3 estimated pitch contour match the tones of the tonal language.
4

5 9. The system of claim 8, further comprising a confidence score augmentor,
6 having a first input coupled to the output of the integrated tone classifier and a second
7 input coupled to the output of the syllable recognition system, for modifying an initial
8 confidence score generated by the syllable recognition system according to a long-term
9 tonal confidence signal generated by the integrated tone classifier to generate an
10 augmented confidence score.
11

12 10. The system of claim 9, wherein the confidence score augmentor
13 comprises:

14 a processor, having inputs and outputs coupled to the inputs and outputs of the
15 long-term tone analyzer, the inputs and outputs of the pitch estimator,
16 and the inputs and outputs of the syllable recognition system, for
17 executing program instruction steps; and

18 a first memory device, having inputs and outputs coupled to the inputs and
19 outputs of the processor, the inputs and outputs of the long-term tone
20 analyzer, the inputs and outputs of the pitch estimator, and the inputs
21 and outputs of the syllable recognition system, for storing program
22 instruction steps for execution by the processor to modify an initial
23 confidence score generated by the syllable recognition system according
24 to a long-term tonal confidence signal generated by the integrated tone
25 classifier to generate an augmented confidence score.
26

27 11. The system of claim 10, wherein the syllable recognition system
28 comprises a second memory device, having inputs and outputs coupled to the inputs
29 and outputs of the processor, the first memory device, and the inputs and outputs of



1 the integrated tone classifier, for storing program instruction steps for execution by the
2 processor to perform short-term tonal analysis of the input signal and for generating
3 N-best theories with initial confidence scores.

4

5 12. The system of claim 11, wherein the integrated tone classifier comprises a
6 third memory device, having inputs and outputs coupled to the inputs and outputs of
7 the processor, the inputs and outputs of the first memory device, and the inputs and
8 outputs of the second memory device, for storing program instruction steps for
9 execution by the processor to receive the N-best theories with initial confidence scores,
10 to perform long-term tonal analysis to determine the tone of syllables of the theories
11 and to generate a long-term tonal confidence signal for each theory.

12

13 13. The system of claim 9, wherein the confidence score augmentor
14 determines an augmented confidence score by adding a long-term confidence score
15 signal and an initial confidence score.

16

17 14. The system of claim 9, wherein the confidence score augmentor re-ranks
18 the N-best theories according to the augmented confidence scores.

19

20 15. A method for recognizing an input signal of continuous speech, having a
21 pitch, of a tonal language, the method comprising the steps of:

22 determining N-best theories with initial confidence scores;

23 estimating the pitch contour of the input signal;

24 segmenting the pitch contour into units according to each of the N-best theories;

25 comparing the units to the tones of the tonal language;

26 generating a long-term confidence signal for each theory, the long-term

27 confidence signal indicating how well the units of a theory match the

28 tones of the tonal language; and



1 modifying an initial confidence score according to a long-term tonal confidence
2 signal to generate an augmented confidence score.

3
4 16. The method of claim 15 further comprising the steps of:
5 ranking the N-best theories according to the augmented confidence scores; and
6 generating an output signal of the N-best theories with augmented initial
7 confidence scores.

8
9 17. A system for recognizing an input signal of continuous speech of a tonal
10 language comprising:

11 means for determining N-best theories with initial confidence scores;
12 means for estimating the pitch contour of the input signal;
13 means for segmenting the pitch contour into units according to each of the N-
14 best theories;
15 means for comparing the units to the tones of the tonal language; and
16 means for modifying an initial confidence score according to a long-term tonal
17 confidence signal to generate an augmented confidence score.

18
19 18. The system of claim 17 further comprising means for generating a long-
20 term tonal confidence signal for each theory.

21
22 19. The system of claim 18 further comprising:
23 means for ranking the N-best theories according to augmented confidence
24 scores; and
25 means for generating an output signal of the N-best theories with augmented
26 confidence scores.

1/5

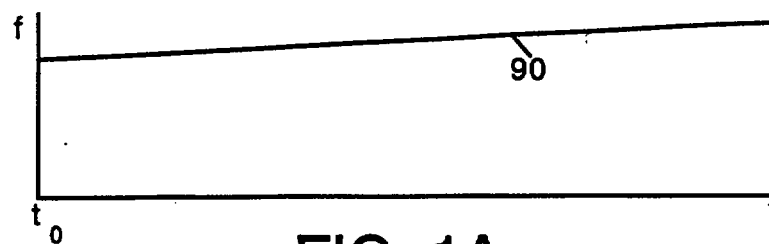


FIG. 1A

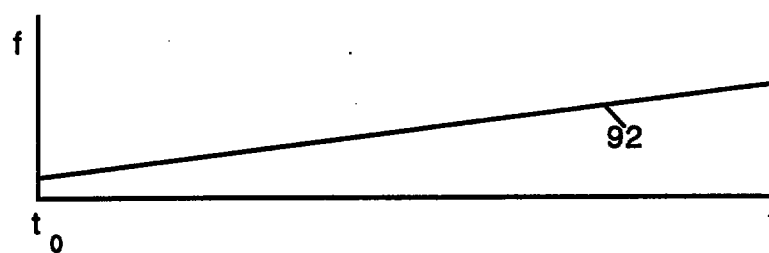


FIG. 1B

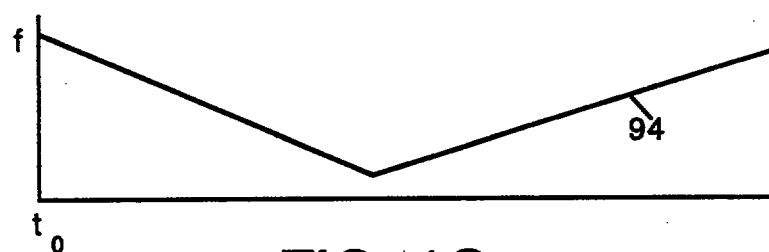


FIG. 1C

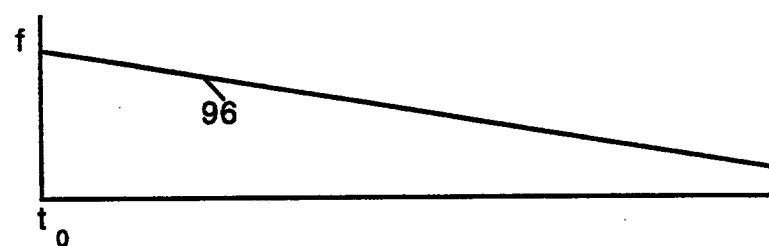


FIG. 1D

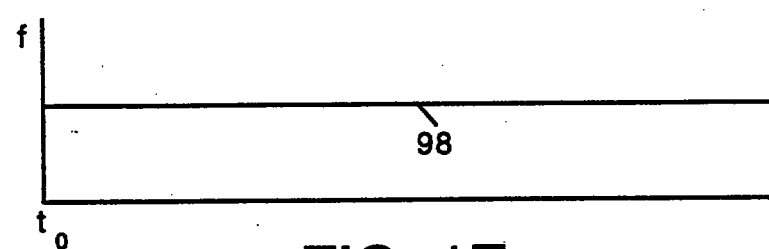


FIG. 1E

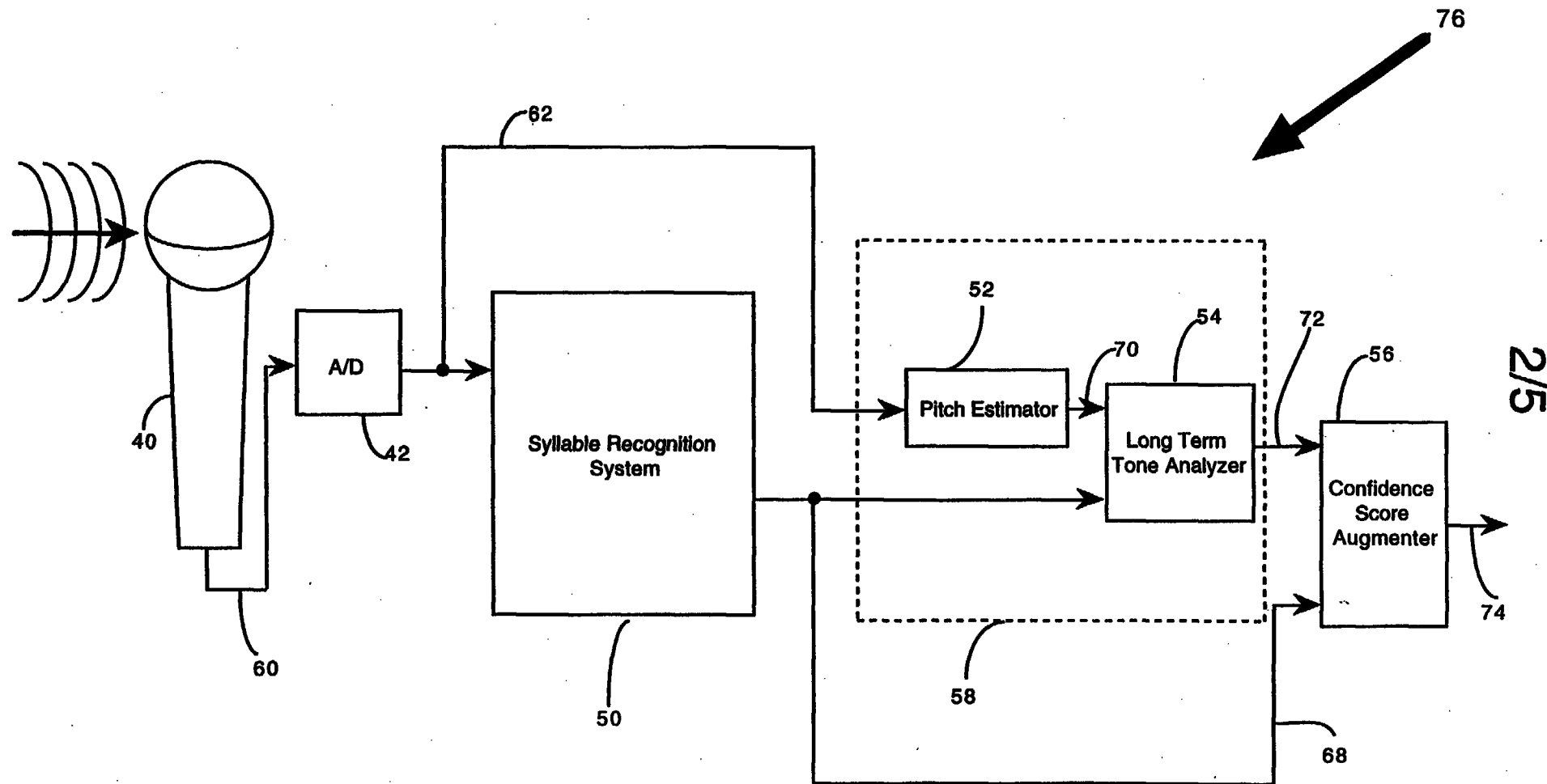


FIG. 2

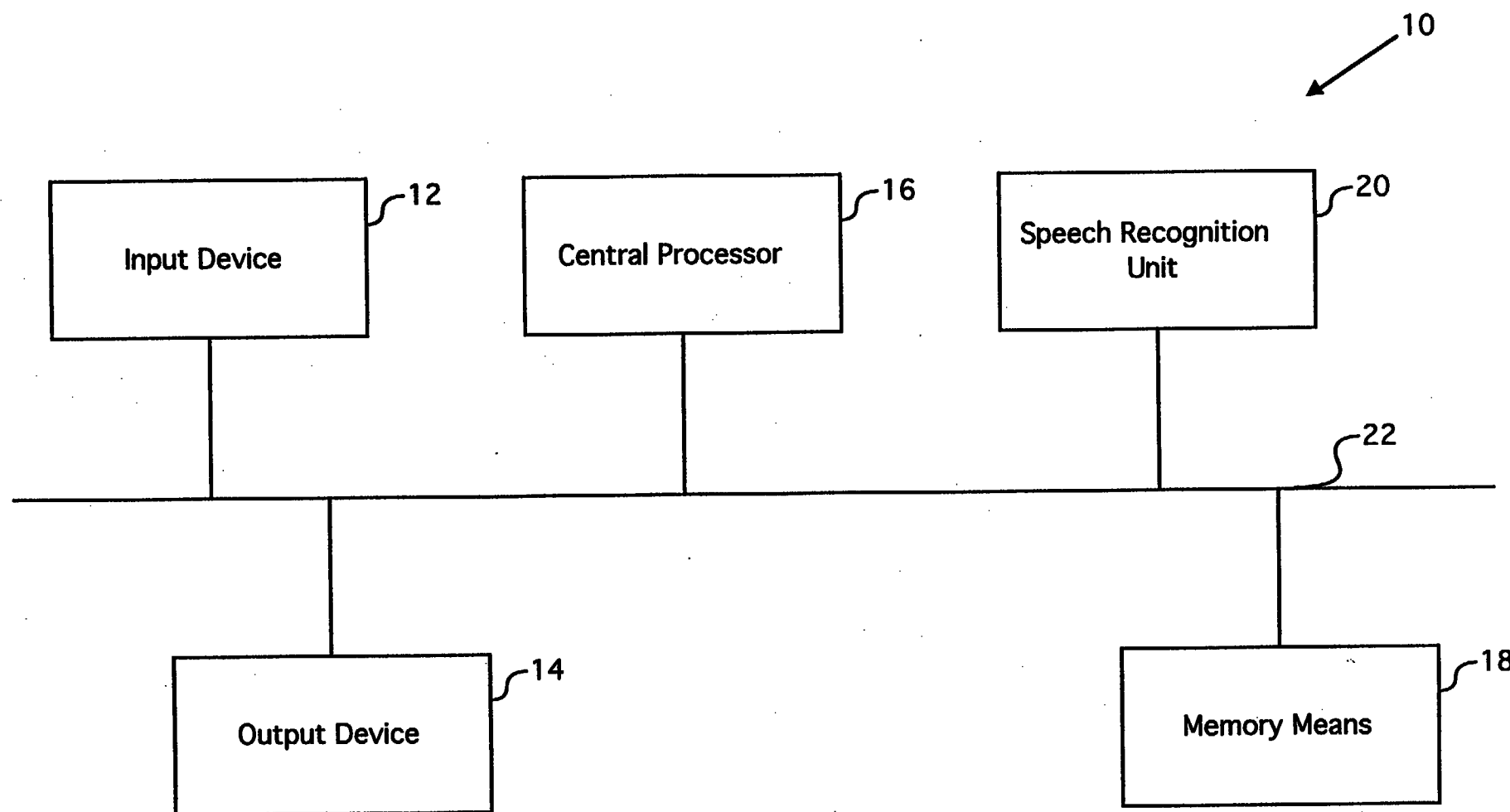


FIG. 3

4/5

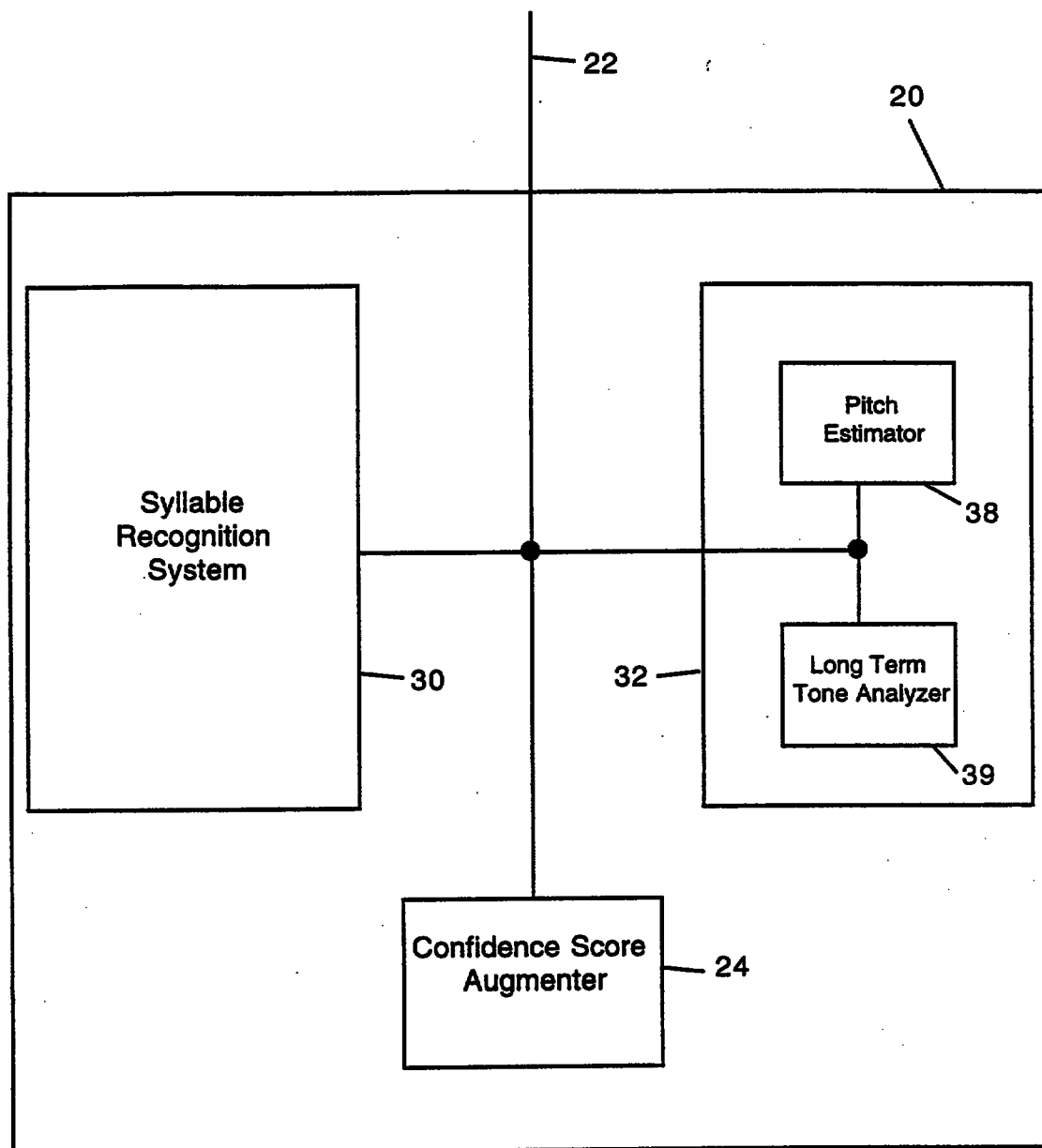


FIG. 4

5/5

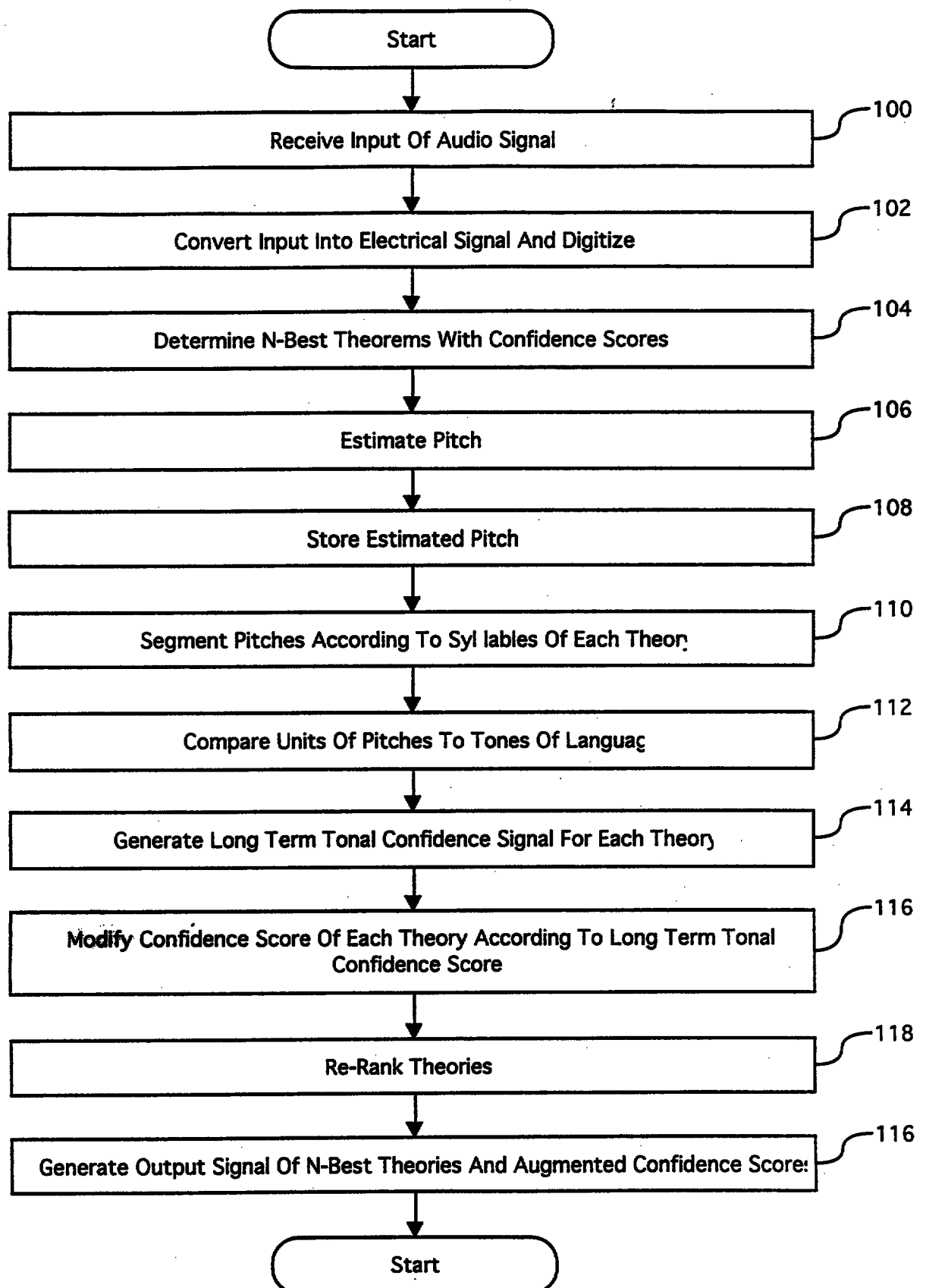


FIG. 5

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International BureauRTmp's
copy

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : G10L 9/00, 9/02, 9/06, 7/08, 5/06, 9/18	A1	(11) International Publication Number: WO 96/10819
		(43) International Publication Date: 11 April 1996 (11.04.96)

(21) International Application Number: PCT/US95/13076

(22) International Filing Date: 29 September 1995 (29.09.95)

(30) Priority Data:
08/316,257 30 September 1994 (30.09.94) US

(71) Applicant: APPLE COMPUTER, INC. [US/US]; One Infinite Loop, M/S 38-1, Cupertino, CA 95014 (US).

(72) Inventors: HON, Hsiao-Wuen; 19445 Dorchester Drive, Saratoga, CA 94070 (US). CHOW, Yen-Lu; 12200 Saraglen Drive, Saratoga, CA 95070 (US). LEE, Kai-Fu; 19120 Austin Way, Saratoga, CA 95070 (US).

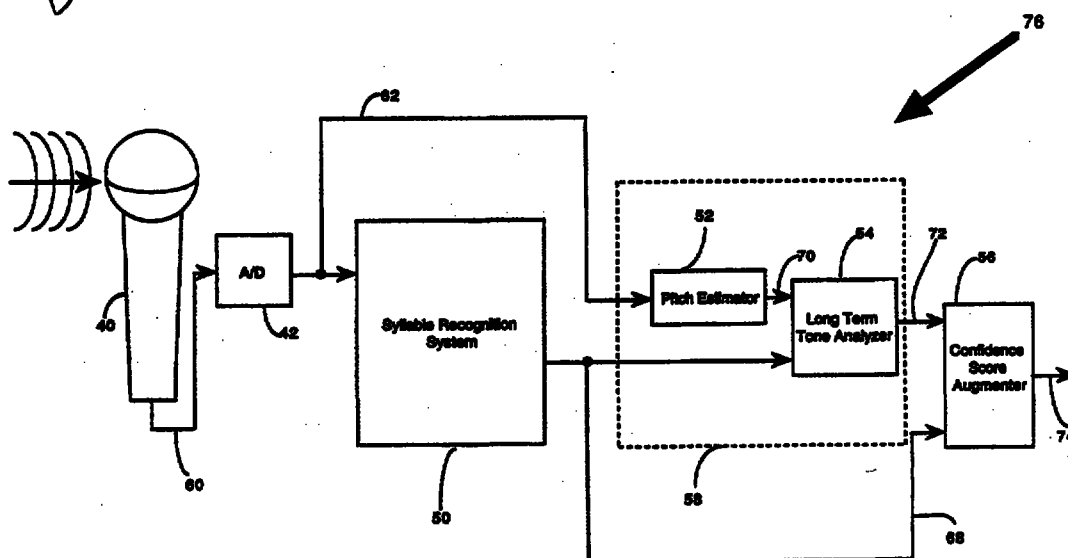
(74) Agents: SUEOKA, Greg, T. et al.; Fenwick & West, Suite 600, Two Palo Alto Square, Palo Alto, CA 94306 (US).

(81) Designated States: AM, AT, AU, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LT, LU, LV, MD, MG, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TT, UA, UG, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG), ARIPO patent (KE, MW, SD, SZ, UG).

Published*With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.*

SG 9701511-9

(54) Title: CONTINUOUS MANDARIN CHINESE SPEECH RECOGNITION SYSTEM HAVING AN INTEGRATED TONE CLASSIFIER



(57) Abstract

A speech recognition system for continuous Mandarin Chinese speech comprises a microphone, and A/D converter, a syllable recognition system, an integrated tone classifier, and a confidence score augmentor. The syllable recognition system generates N-best theories with initial confidence scores. The integrated tone classifier has a pitch estimator to estimate the pitch of the input once and a long-term tone analyzer to segment the estimated pitch according to the syllables of each of the N-best theories. The long-term tone analyzer performs long-term tonal analysis on the segmented, estimated pitch and generates a long-term tonal confidence signal. The confidence score augmentor receives the initial confidence scores and the long-term tonal confidence signals, modifies each initial confidence score according to the corresponding long-term tonal confidence signal, re-ranks the N-best theories according to the augmented confidence scores, and outputs the N-best theories.

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.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgyzstan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LU	Luxembourg	TD	Chad
CS	Czechoslovakia	LV	Latvia	TG	Togo
CZ	Czech Republic	MC	Monaco	TJ	Tajikistan
DE	Germany	MD	Republic of Moldova	TT	Trinidad and Tobago
DK	Denmark	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	US	United States of America
FI	Finland	MN	Mongolia	UZ	Uzbekistan
FR	France			VN	Viet Nam
GA	Gabon				

970 15 11 -9