The present invention relates to a continuous speech recognizing apparatus for carrying out speech recognition using a probabilistic language model and to a recording medium.

An apparatus for recognizing continuous speech using a multiple-pass decoder is one of the speech recognition apparatuses using a probabilistic language model. The second pass circuit of this apparatus cannot determine the speech recognition result until the speech recognition candidates of a sentence have been generated. This causes a large time lag (delay) between the input instant of the speech and the output of the speech recognition result from the continuous speech recognizing apparatus. Therefore, the object the present invention is to provide a continuous speech recognizing apparatus and recording medium capable of reducing the time lag between the input of speech and the output of the speech recognition result of a multiple-pass speech recognizing apparatus.

Figure 1 shows the functional configuration of a continuous speech recognizing apparatus in accordance with the present invention. An acoustic analyzer 1 carries out A/D conversion, followed by acoustic analysis, and outputs parameters indicating speech features. A first-pass processor 2 uses a simple probabilistic language model, and successively generates a word lattice 4, which is composed of the end time of each word candidate, its score, and a pointer to the previous word. Then, it traces back the word lattice 4 beginning from the phoneme with the current maximum score for every $\Delta t$ frames of the input speech, thereby obtaining word string candidates (N-best word strings).

A second-pass processor 3 rescoring the N-best word strings generated every $\Delta t$ frames by using a more complex probabilistic language model 6, and selects from among the N-best word strings the best word string with the maximum score. The second pass processor 3 compares the current 1-best word string with the previous 1-best word string obtained for the previous frames to detect a stable portion in the current 1-best word string, and detect when word string is the same as in the speech recognition result (see, FIG. 2).

As described above, according to the present invention, the stable portion in the word string (the 1-best word string in the embodiment) having the maximum likelihood is detected by executing the successive two-pass processing, and becomes the partial speech recognition results.