SYSTEM AND METHOD FOR PROCESSING CHINESE LANGUAGE TEXT

ABSTRACT

Phonetic Chinese is entered into a computer system and accurately converted into the Hanzi form. The system has a novel keyboard (1030) with diacritic keys (1050) that permit the user to annotate each entered phonetic text syllable with a diacritic that indicates the tone of the syllable. A process executing on the system determines that a syllable has been entered when a diacritic key (1050) is struck. An entered phonetic syllable is then compared to a list of acceptable phonetic syllables and abbreviations. If the entered syllable is on the list, the correctly spelled and accented syllable is stored in memory and displayed on a phonetic portion of a graphical display (1020). The process continues for succeeding syllables until a delimiter is entered. Upon encountering a delimiter, the word string is analyzed using morphological and syntactical processes and/or a statistical language model to unambiguously determine the proper Hanzi characters that represent the word(s) in the word string.

Fig. 1