METHODS AND SYSTEMS OF ROUTING UTTERANCES BASED ON CONFIDENCE ESTIMATES

CROSS REFERENCE TO RELATED APPLICATIONS

This application claims benefit of U.S. Provisional Application No. 60/210,823, filed Jun. 12, 2000, which is incorporated by reference.

TECHNICAL FIELD

This invention relates to confidence estimates in speech recognition.

BACKGROUND

The computation of word-level confidence estimates for the results produced by a speech recognition system has become a well-established practice. In general, each of these estimates indicates the probability that a particular word in the results is correct (that is, that the speaker actually said the word). Word-level confidence estimates have been applied to such tasks as spotting misrecognized or out-of-vocabulary words, rejecting recognition hypotheses in command-and-control environments, and controlling prompts for confirmation in computer-based dialogue systems. Techniques for determining word-level confidence estimates are described by L. Gillick, Y. Ito, and J. Young in "A Probabilistic Approach to Confidence Estimation and Evaluation," Proc. ICASSP-97, pages 879–882 (1997), which is incorporated by reference.

SUMMARY

In one general aspect, routing an utterance to a system includes receiving an utterance, and processing the utterance using large-vocabulary continuous speech recognition to generate a string of text corresponding to the utterance. A confidence estimate of the string of text corresponding to the utterance is generated and compared to a predetermined threshold. If the confidence estimate satisfies the predetermined threshold, the string of text is forwarded to the system. If the confidence estimate does not satisfy the predetermined threshold, the information relating to the utterance is forwarded to a transcriptionist.

Implementations may include one or more of the following features. For example, the utterance may include two or more words. The transcriptionist may determine an acceptable string of text and forwarding the acceptable string of text to the system if the confidence estimate does not satisfy the predetermined threshold.

The confidence estimate may indicate a probability that the string of text is an acceptable representation of the utterance. The information relating to the utterance may include one or more of: the utterance, the string of text corresponding to the utterance, and the generated confidence estimate of the string of text.

The confidence estimate may be generated by selecting one or more predictors relating to the large-vocabulary continuous speech recognition, and by training a confidence model using the one or more predictors. The confidence estimate may be generated by extracting values of the one or more predictors based on the received utterance, and by providing the extracted values to the confidence model to generate the confidence estimate.

The confidence estimate may be further compared to a second predetermined threshold. If the confidence estimate does not satisfy the first predetermined threshold and does satisfy the second predetermined threshold level, the information relating to the utterance may be forwarded to the user who spoke the utterance, and the user may be permitted to act in response to the forwarded information.

The transcriptionist may be a human transcriptionist, and the system may include a human recipient.

In another general aspect, routing a message to a system includes receiving a message including utterances, and processing each utterance in the message using large-vocabulary continuous speech recognition to generate a string of text corresponding to that utterance. A confidence estimate is generated for each string of text that corresponds to an utterance and each confidence estimate is compared to a predetermined threshold. If all of the confidence estimates satisfy the predetermined threshold, the string of text is forwarded to the system. If any one of the confidence estimates does not satisfy the predetermined threshold level, the message is forwarded to a transcriptionist.

Implementations may include one or more of the following features. For example, the confidence estimate for a string of text may indicate a probability that the string of text is an acceptable representation of the corresponding utterance.

The transcriptionist may determine acceptable strings of text for the message and forward the acceptable strings of text for the message to the system if any one of the confidence estimates does not satisfy the predetermined threshold.

Generation of the confidence estimate for a string of text may include selecting one or more predictors for the string of text based on the large-vocabulary continuous speech recognition and training a confidence model for the string of text using the one or more predictors. Generation of the confidence estimate for a string of text may include extracting values of the one or more predictors for the string of text based on the corresponding utterance and providing the extracted values for the utterance to the confidence model to generate the confidence estimate for the string of text.

Each confidence estimate may be compared to a second predetermined threshold. In this case, if a confidence estimate does not satisfy the first predetermined threshold level and satisfies the second predetermined threshold level, information relating to the message is forwarded to the user who spoke the message, and the user is permitted to act in response to the forwarded information. The information relating to the message may include one or more of the message, the string of text for the confidence estimate, and the confidence estimate for the string of text.

The transcriptionist may be a human transcriptionist, and the system may include a human recipient.

In another general aspect, routing a message to a system includes receiving a message including utterances and processing each utterance in the message using large-vocabulary continuous speech recognition to generate a string of text corresponding to that utterance. A confidence estimate is generated for each string of text that corresponds to an utterance and each confidence estimate is compared to a predetermined threshold. If all of the confidence estimates satisfy the predetermined threshold, the strings of text are forwarded to the system. If one of the confidence estimates does not satisfy the predetermined threshold, information relating to the utterance corresponding to that confidence estimate is forwarded to a transcriptionist.