

Abstract

A speech recognition method, comprising: acquiring speech data to be recognized (302); extracting a Filter Bank feature and a Mel-Frequency Cepstral Coefficient (MFCC) feature from the speech data (304); using the MFCC feature as input data of a trained Gaussian Mixture Model-Hidden Markov Model (GMM-HMM) model, acquiring a first likelihood probability matrix outputted by the trained GMM-HMM model (306); using the Filter Bank feature as an input feature of a trained long short-term memory (LSTM) model which has a connection unit, acquiring a posterior probability matrix outputted by the LSTM model which has the connection unit, the connection unit being used for controlling the flow of information between layers in the LSTM model (308); using the posterior probability matrix and the first likelihood probability matrix as input data of a trained HMM model, acquiring an outputted second likelihood probability matrix (310); and acquiring a corresponding target word sequence corresponding to the speech data to be recognized in a phoneme decoding network according to the second likelihood probability matrix (312).