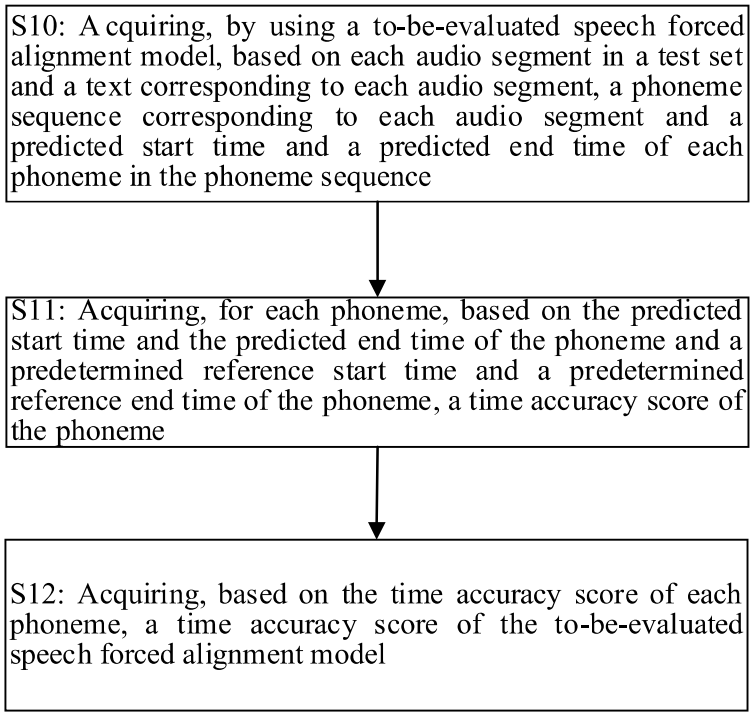

S10: Acquiring, by using a to-be-evaluated speech forced alignment model, based on each audio segment in a test set and a text corresponding to each audio segment, a phoneme sequence corresponding to each audio segment and a predicted start time and a predicted end time of each phoneme in the phoneme sequence
S11: Acquiring, for each phoneme, based on the predicted start time and the predicted end time of the phoneme and a predetermined reference start time and a predetermined reference end time of the phoneme, a time accuracy score of the phoneme
S12: Acquiring, based on the time accuracy score of each phoneme, a time accuracy score of the to-be-evaluated speech forced alignment model

Fig. 1

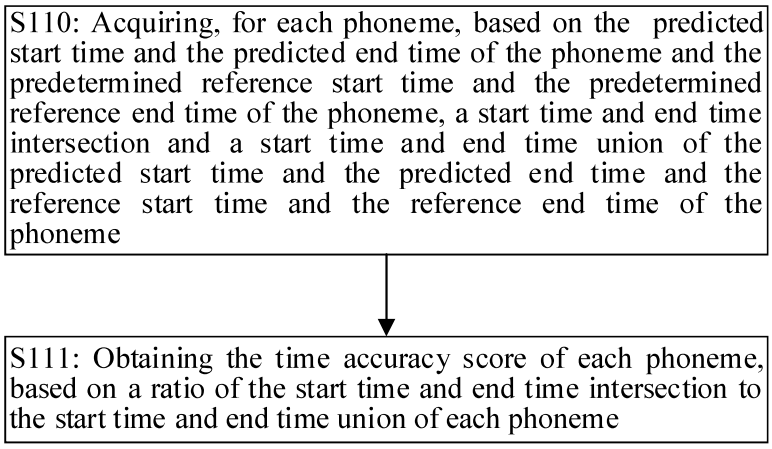

S110: Acquiring, for each phoneme, based on the predicted start time and the predicted end time of the phoneme and the predetermined reference start time and the predetermined reference end time of the phoneme, a start time and end time intersection and a start time and end time union of the predicted start time and the predicted end time and the reference start time and the reference end time of the phoneme
S111: Obtaining the time accuracy score of each phoneme, based on a ratio of the start time and end time intersection to the start time and end time union of each phoneme

Fig. 2

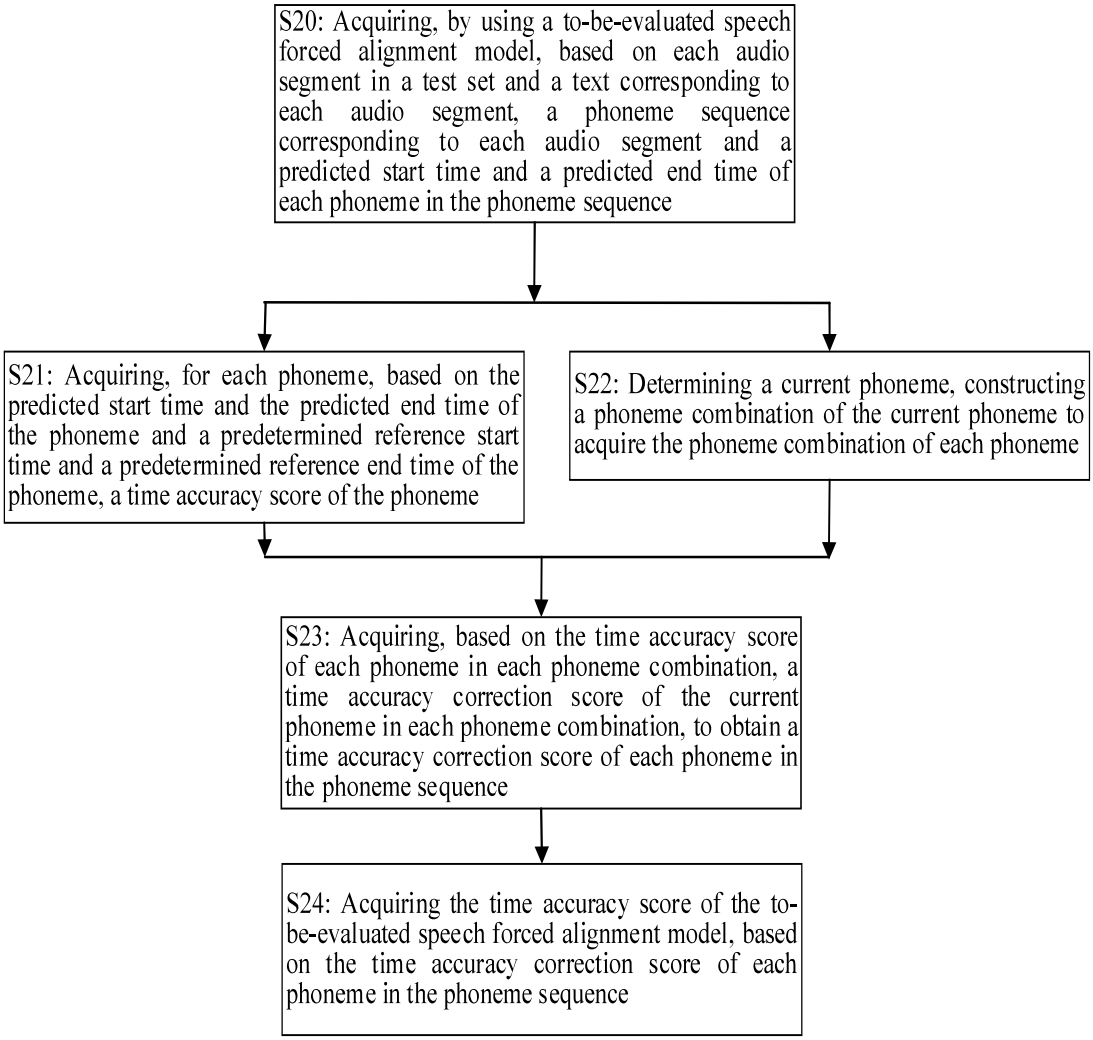
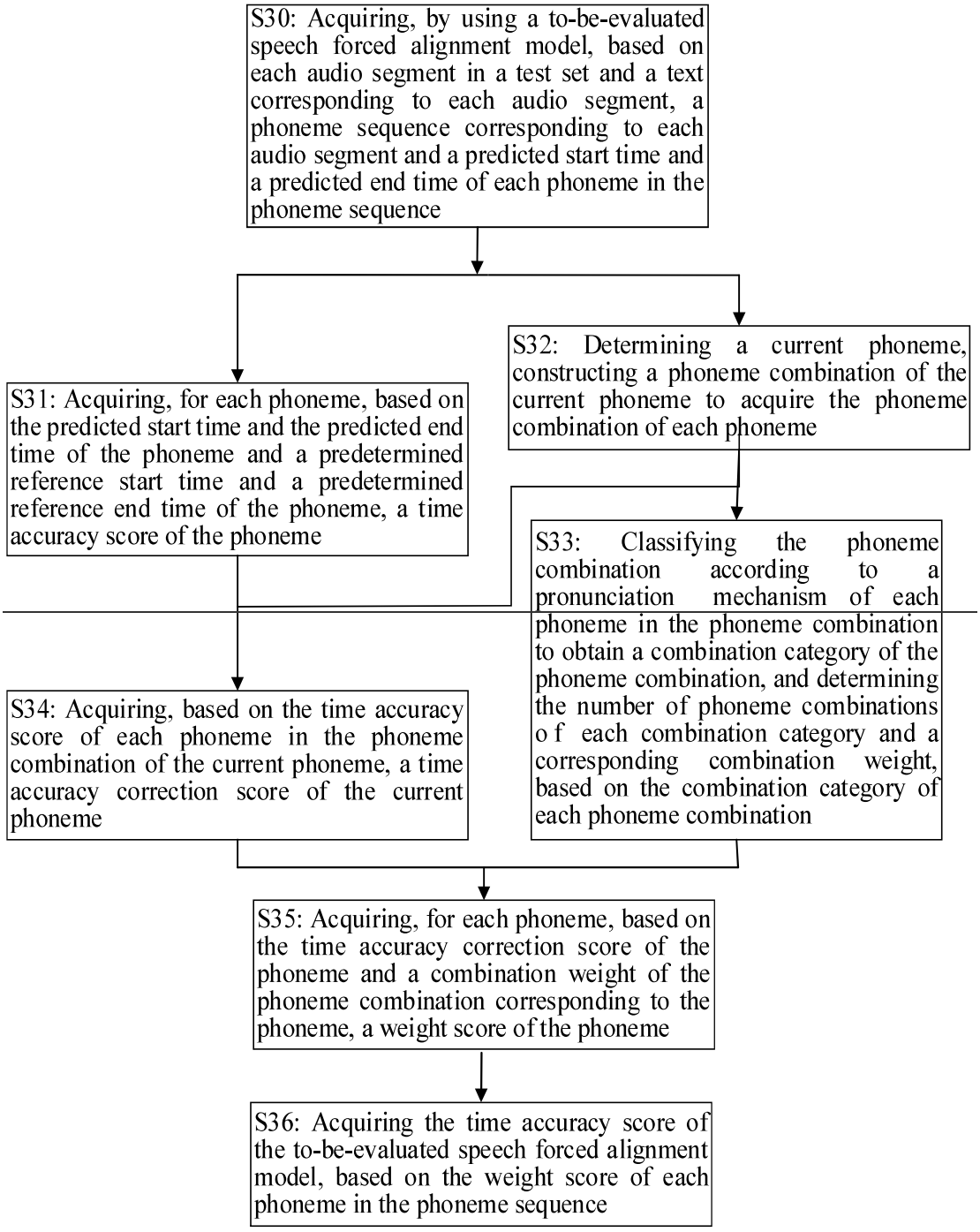

S20: Acquiring, by using a to-be-evaluated speech forced alignment model, based on each audio segment in a test set and a text corresponding to each audio segment, a phoneme sequence corresponding to each audio segment and a predicted start time and a predicted end time of each phoneme in the phoneme sequence
S21: Acquiring, for each phoneme, based on the predicted start time and the predicted end time of the phoneme and a predetermined reference start time and a predetermined reference end time of the phoneme, a time accuracy score of the phoneme
S22: Determining a current phoneme, constructing a phoneme combination of the current phoneme to acquire the phoneme combination of each phoneme
S23: Acquiring, based on the time accuracy score of each phoneme in each phoneme combination, a time accuracy correction score of the current phoneme in each phoneme combination, to obtain a time accuracy correction score of each phoneme in the phoneme sequence
S24: Acquiring the time accuracy score of the to-be-evaluated speech forced alignment model, based on the time accuracy correction score of each phoneme in the phoneme sequence

Fig. 3


S30: Acquiring, by using a to-be-evaluated speech forced alignment model, based on each audio segment in a test set and a text corresponding to each audio segment, a phoneme sequence corresponding to each audio segment and a predicted start time and a predicted end time of each phoneme in the phoneme sequence
S31: Acquiring, for each phoneme, based on the predicted start time and the predicted end time of the phoneme and a predetermined reference start time and a predetermined reference end time of the phoneme, a time accuracy score of the phoneme
S32: Determining a current phoneme, constructing a phoneme combination of the current phoneme to acquire the phoneme combination of each phoneme
S33: Classifying the phoneme combination according to a pronunciation mechanism of each phoneme in the phoneme combination to obtain a combination category of the phoneme combination, and determining the number of phoneme combinations o f each combination category and a corresponding combination weight, based on the combination category of each phoneme combination
S34: Acquiring, based on the time accuracy score of each phoneme in the phoneme combination of the current phoneme, a time accuracy correction score of the current phoneme
S35: Acquiring, for each phoneme, based on the time accuracy correction score of the phoneme and a combination weight of the phoneme combination corresponding to the phoneme, a weight score of the phoneme
S36: Acquiring the time accuracy score of the to-be-evaluated speech forced alignment model, based on the weight score of each phoneme in the phoneme sequence

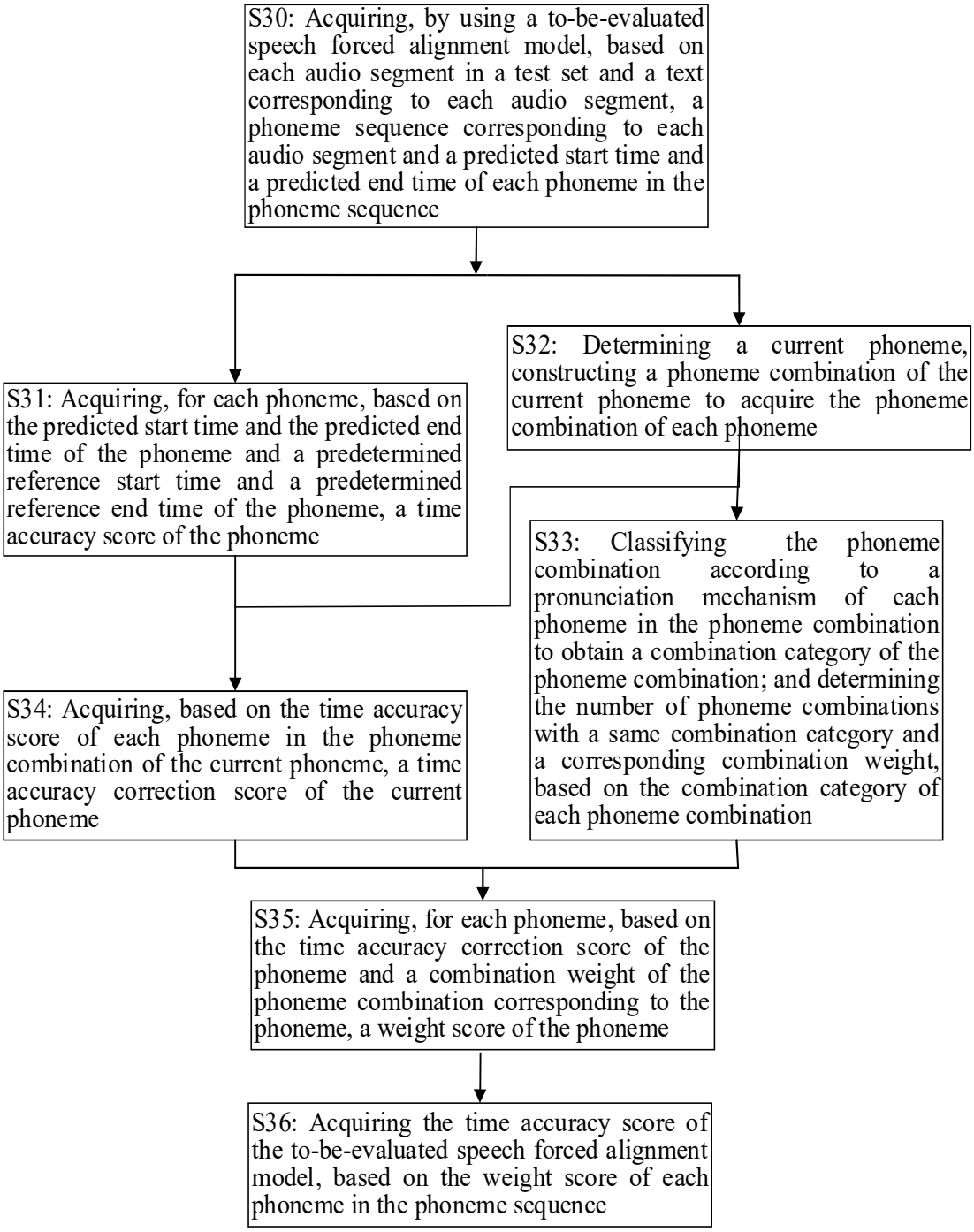

S30: Acquiring, by using a to-be-evaluated speech forced alignment model, based on each audio segment in a test set and a text corresponding to each audio segment, a phoneme sequence corresponding to each audio segment and a predicted start time and a predicted end time of each phoneme in the phoneme sequence
S31: Acquiring, for each phoneme, based on the predicted start time and the predicted end time of the phoneme and a predetermined reference start time and a predetermined reference end time of the phoneme, a time accuracy score of the phoneme
S32: Determining a current phoneme, constructing a phoneme combination of the current phoneme to acquire the phoneme combination of each phoneme
S33: Classifying the phoneme combination according to a pronunciation mechanism of each phoneme in the phoneme combination to obtain a combination category of the phoneme combination; and determining the number of phoneme combinations with a same combination category and a corresponding combination weight, based on the combination category of each phoneme combination
S34: Acquiring, based on the time accuracy score of each phoneme in the phoneme combination of the current phoneme, a time accuracy correction score of the current phoneme
S35: Acquiring, for each phoneme, based on the time accuracy correction score of the phoneme and a combination weight of the phoneme combination corresponding to the phoneme, a weight score of the phoneme
S36: Acquiring the time accuracy score of the to-be-evaluated speech forced alignment model, based on the weight score of each phoneme in the phoneme sequence

Fig. 4

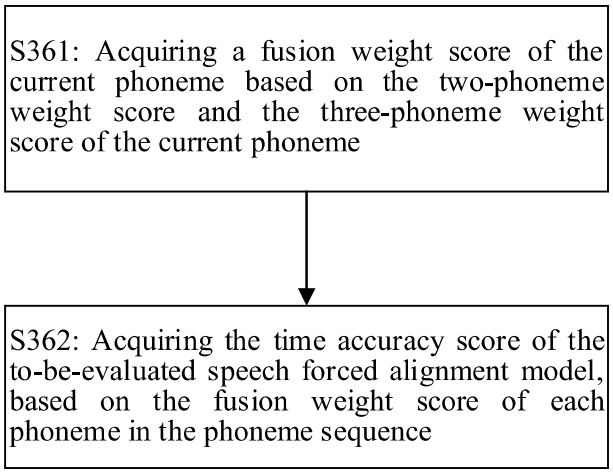

S361: Acquiring a fusion weight score of the current phoneme based on the two-phoneme weight score and the three-phoneme weight score of the current phoneme
S362: Acquiring the time accuracy score of the to-be-evaluated speech forced alignment model, based on the fusion weight score of each phoneme in the phoneme sequence

Fig. 5

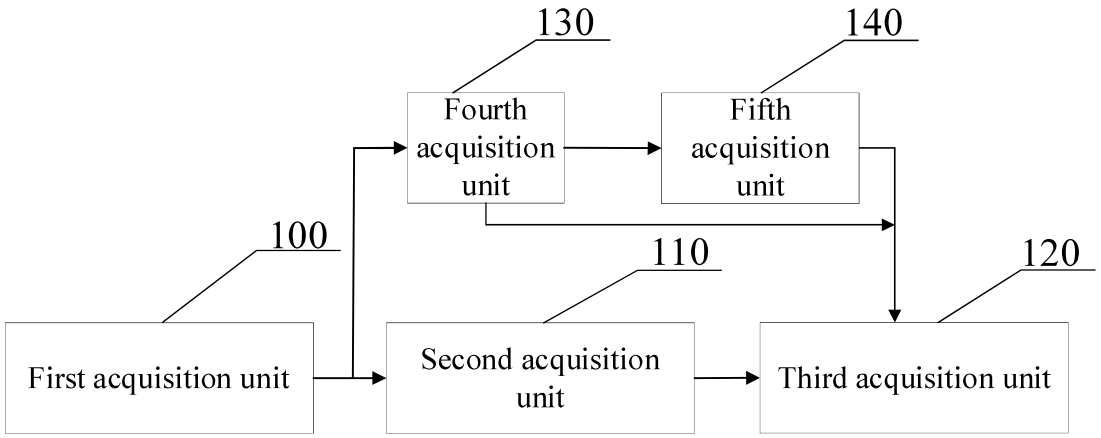

130
140
Fourth acquisition unit
Fifth acquisition unit
100
110
120
First acquisition unit
Second acquisition unit
Third acquisition unit

Fig. 6

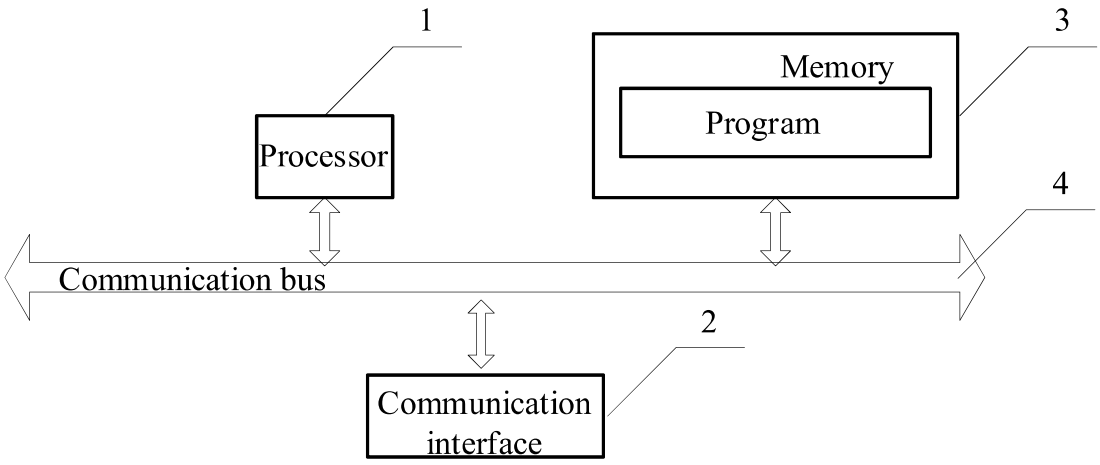

1
3
Memory
Program
Processor
4
Communication bus
2
Communication interface

Fig. 7