* The system conducts searching database and answers to the user in speech.

* When finishing, it returns to the step (0)

_ (4.1) If it is not the case, which means that the command is grammatically incorrect, the system will ask the user to redo a command in speech.

We will follow the example of a specific transaction:

- The system: Hello.
- The user: Please tell me what the weather is like in Can Tho tomorrow?
- The system: Do you want to know about the weather of tomorrow in Can Tho?
- The user: That's right.
- -The system: The temperature is between 24 degrees and 34 tomorrow, it is sunny in the daytime, it rains in the evening and at night.

For achieving the above functions, the system must require the following components:

- Voice recognizer component: to transform speech data which is human speech into data text.
- Vietnamese language processing component: to analyze commands' syntax and semantic meaning from users.
- Central processing component: to connect the other components each to another through such operations:

1. Transforming data text from speech recognizer into standard data to be executed by Prolog in Vietnamese language processing.

2. Transforming commands' semantic expressions into a file of select statements which is sent to database and executing them.

3. Filtering, arranging, and returning processed results via system to user.

- Database: contains selected data.
- Vietnamese language synthetizing component: transforms text data into speech.

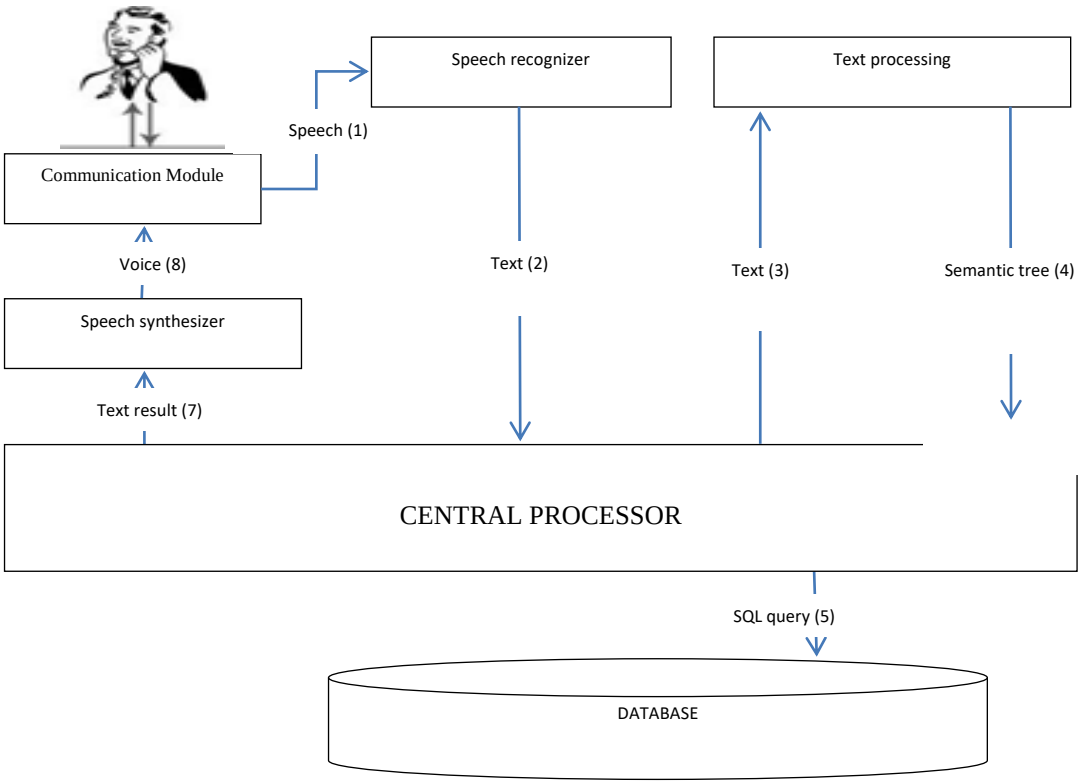

Figure 1: system structure.

3. Voice recognizer component

In the Weather Voice system, we use HTK to build the voice recognizer component. HTK provides us with instruments analyzing voice, specially the one recognizing voice, based on HMM [13]. According to approaches of [9], [10], [15], [16], [18] we implement a context-dependent model making use of triphone to recognize words from word list as well as to identify patterns' grammatical meaning which can happen in application context for the sake of recognizing sentences more correctly.

A. Steps in building voice recognizer component.

Building a voice recognizer component includes two main periods:

1. Training period:

- Preparing voice data file needs training and codifying this file.

- Labelling, building dictionary.
- Creating HMM prototypes for each of phone units.

The training period's output is the file of trained HMM prototypes.

2. Recognizing period:

- The file of trained HMM set is the result of the training period.
- Building dictionary.
- Extracting characteristics for voice series needing being recognized.

The recognizing period's output is the text series.

B. Data to train.

The file of data to train is recorded for 200 minutes in amount of 2,500 patterns. Those data must satisfy the criterion of 8,000 Hz, 16bit according to PCM format and be recorded in 50 different accents and in a quiet environment. The word list comprises 63 cities and towns' names over the country and key words concerning command sentences.

C. Building language grammar.

Our grammar is language prototypes providing information of sentences' syntax, semantics and word order. This component will help the system choose the best recognized results from the list of 'candidates' previously selected by the recognizing period.

Sentence structures are likely to exist in application contexts.

Building a language prototypes involves the determining its grammar. The complexity of grammar depends on that of the system needing being recognized.

Grammar structure is a generalized graphic which implicates pattern sentences possibly occuring in application context. In our application, a part of the grammar file will be displayed as:

```
$tinh_thanh= HOOF CHIS MINH | HAF NOOJI | DDAF NAWXNG | CAAFN THOW;

$ngay=HOOM NAY | NGAFY MAI | NGAFY MOOST;

…

$sentence1=THOWFI TIEEST $tinh_thanh $ngay(NHUW THEES NAFO | RA SAO);
```

D. Voice synthesizing (text-to-speech).

Text-to-speech system includes two main phases: text analyzing (processing phase, standardizing text input in order to synthesize it) and text -to - speech phase (building speech signals from the former's results). The second phrase can be implemented by Formant text-to-speech [8] or Unit-selection approach [1],[8]… For Weather Voice, we have chosen unit-selection, the phrase is undergone in the next procedure:

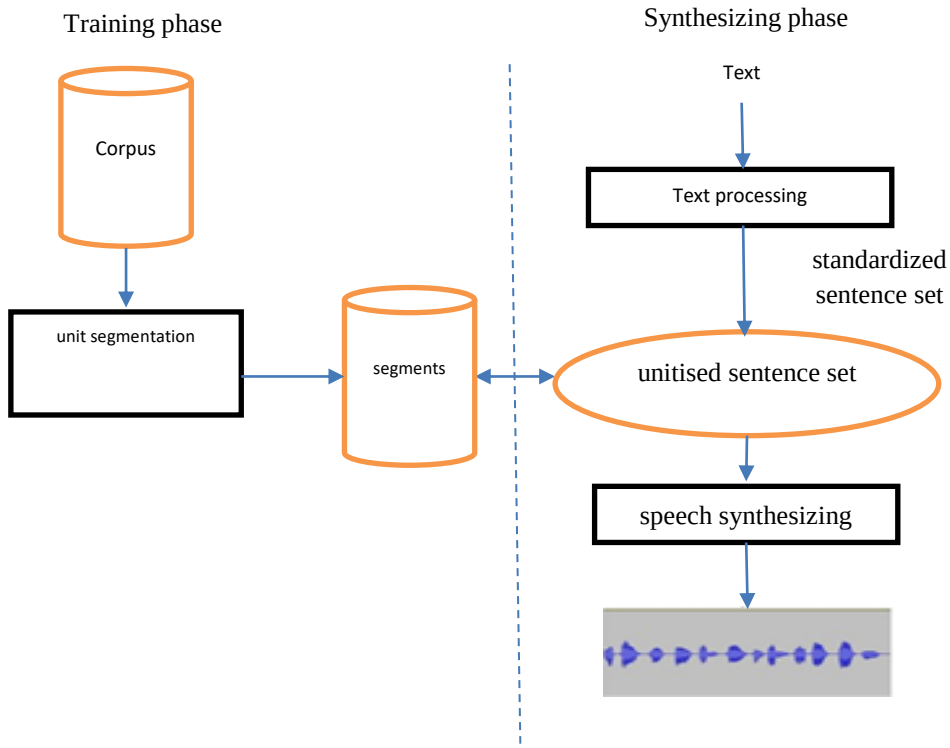

Figure 6: Text-to-speech procedure by connecting to unitise.

4. Vietnamese language processing

A. Query statement syntax.

Our system in all represents 18 query statement patterns as shown in the table 1.

Table 1: Some query statements classified according to them.

Number	Query statement pattern	Example
1	[weather] [place] [time]	What is the weather like in Hanoi today?
2	[weather] [place]	What is the weather like in Hanoi?
3	[weather] [time]	What is the weather like today?
4	[place] [time] [weather]	In Hanoi tomorrow, what is the weather like?
5	[place] [weather]	In HCM city what is the weather like?
6	[time] [weather]	After tomorrow what is the temperature?
7	[time] [place] [weather]	After tomorrow in Can Tho what is the weather like?
8	[temperature] [place] [time]	What is the temperature in Hanoi today?
9	[temperature] [place]	What is the temperature in Hanoi?
10	[temperature] [time]	What is the temperature today?
11	[place] [temperature] [time]	In Hanoi what is the temperature tomorrow?
12	[place] [temperature]	In HCM city what is the temperature?
13	[time] [temperature]	After tomorrow what is the temperature?
14	[time] [place] [temperature]	After tomorrow in Can Tho what is the temperature?
15	[place] [time] [state]	In Da Nang tomorrow does it rain?
16	[time] [place] [state]	Tomorrow in Da Nang is it hot?
17	[place] [state]	In Sai Gon does it rain?
18	[time] [state]	After tomorrow does it rain?

B. Analyzing statement semantic.

To represent query statement semantic, we use DCG [3],[7],[11],[12],[17] all nine representation patterns of query statement semantic are given in table 2:

Table 2. Representation patterns of query statement semantic.

Number	**Meaning representations**	**Query statement patterns**
1	query(weather, place, time)	Structure patterns from 1 to 7 coming from table 2
2	query (weather, place)	
3	query (weather, time)	
4	query (temperature, place, time)	Structure patterns from 8 to 14 coming from table 2
5	query (temperature, place)	
6	query (temperature, time)	
7	yesno (place, state, time)	Structure patterns from 15 to 18 coming from table 2
8	yesno (place, state)	
9	yesno (state, time)	

Example 1: What is the weather like in Hanoi tomorrow?

DCG syntactic and semantic order will be defined as follows:

query(query(Weather,Place,Time)) --> n_weather(Weather),prep_place,n_place(Place), n_time(Time),w_how.

prep_place --> [in].

n_weather(weather) --> [weather].

n_place(place(cần_thơ)) --> [cần_thơ].

n_time(time(tomorrow)--> [tomorrow].

w_how --> [how].

DCG syntactic and semantic order has defined semantic pattern of example 1's query statement as:

query(weather(thoi_tiet), place(can_tho), time(ngay_mai)).

This pattern is the pattern number 1 in the table 3.

We carry out transforming these semantic patterns into corresponding SQL statements in order to search corresponding weather information in the database. These data are automatically extracted from Yahoo Weather APL.

5. Experiments and evaluations

The experiments, at first, were conducted by each of system's components, which comprised voice recognizer, Vietnamese language processor and central processor. Afterwards, we experimented the entire system as well as performed surveys on users' reviews of the system including text-to-speech component.

A. Speech recognizer component.

The performance of speech recognizer is often evaluated by the metric WER (Word Error Rate), it is computed as the next formula:

$$WER= (S+D+I) / N \times 100\%$$

Where:

- N is the number of words,
- S is the number of substitutions,
- I is the number of insertions,
- D is the number of deletions.

But here, we used the metric WAR (Word Accuracy Rate) in the evaluation of the system's performance by the formula:

$$WAR= (1– (S+ D+ I) / N) \times 100\%$$

We gradually carried out the experiments classified in accordance to areas, sexes, ages, and trained participants. The system's accuracy is shown in the following tables 3, 4, 5 and 6.

Table 3: Based on areas.

WAR		
North	Center	South
90%	88%	93%

Table 4: Based on sexes.

WAR	
Female	**Male**
91%	94%

Table 5: Based on ages.

WAR	
18-30	Others
95%	90%

Table 6: Based on trained participants and non-participants.

WAR	
Trained participants	Non-participants
98%	93%

B. Natural language processing component.

Thanks to this component, we succeeded in experimenting 50 statements, the results for the statements involved exactly corresponded to our expectation. These statements are circumscribed within the syntastic structures having been built for the system. The system's ability to correctly process all these standard statements attests its stability and accuracy.

The amplitude: For the statements which do not fall within the syntastic structures' circumscription, when returned by the system they will be assessed to be false. This evidence

demonstrates our synstatic regulations DCG is incomplete and yet our grammar cannot cover every case which is likely to occur.

If words grammar is added to it and synstatic regulations are improved, the system's amplitude will significantly expand.

C. Surveys on users' reviews.

During these surveys, users were asked the question coming from the system: "Is the system easy to use?". The answers are ranked in order of usefulness as mentioned in the table 7.

Table 7: Indication of levels of the system's usefulness.

Very useful	Quite useful	Slightly useful	Useless
25%	27%	23%	25%

D. Experimenting the entire system.

The system is built in PC environment with Java language and SWI-Prolog version 6.6.5.

Table 8: Experiment parameter.

Number of query statements	100
Environment	indoor
Sample rate	8kHz
Quantization	16bits
Format	PCM
Device	mobile phone

The results given by the system are correct for 48 out of 50 Vietnamese query statements. As seen above, all the unexpected results happened during the identifying period. The feedback time is 2.6 seconds on average.

E. Assessment

Throughout the experiment, voice recognizer component has incorrectly identified 6 out of 50 experimented statements. But syntactically, among them 4 statements have always kept their initial meanings and then correctly been processed by natural language processing component; only the two left ones have received the wrong meanings because of the identifying period. By that, we have found that the important part is played by the natural language processing component, it can even rectify errors resulting from the identifying period.

6. Conclusion

The paper has offered the presentation of the Weather Voice system's structure and the approaches to build it. In the system, the Vietnamese language processing component, responsible for analyzing the statements' syntax and semantic meaning, assumes the central part. According to our knowledge, this is one of the first systems in Vietnam to be equipped with effective natural language processing mechanism of voice application, this priority has made the system more intelligent and flexible. This research has also opened a new development to building and expanding question-answer systems which can understand Vietnamese and communicate with users in this language. In the future, we are also interested in developing voice applications for emotional analysis based on studies by Thien et al. [19,20].

7. References

[1] Hunt, A. Black and W. Alan, “Unit selection in a concatenative speech synthesis system using a large speech database," Pro c. ICASSP-96, 1, pp. 373, 1996.

[2] Duong Dau, Minh Le, Cuong Le and Quan Vu, “A Robust Vietnamese Voice Server for Automated Directory Assistance Application,” RIVF-VLSP 2012, Ho Chi Minh City, Viet Nam, 2012.

[3] Fernando C. N. Pereira and Stuart M. Shieber, Prolog and Natural-Language Analysis. Microtome Publishing, pp.1 – 284, Massachusetts, 2005.

[4] Hue Nguyen, Truong Tran, Nhi Le, Nhut Pham, Quan Vu, “iSago: The Vietnamese Mobile Speech Assistant for Food-court and Restaurant Location,” RIVF-VLSP 2012, Ho Chi Minh City, Viet Nam, 2012.

[5] Michelle Quinton, Windows NT 5.0 Brings You New Telephony Development Features with TAPI 3.0, Microsoft Systems Journal. [Online]. Available: http://www.microsoft.com/msj/1198/tapi3/tapi3.aspx, 1998.

[6] Nhut Pham, Quan Vu, "A Spoken Dialog System for Stock Information Inquiry," in Proc. IT@EDU, Ho Chi Minh City, Viet Nam, 2012.

[7] Patrick Blackburn, Johan Bos, "Representation and Inference for Natural Language: A First Course in Computational Semantics". CSLI Press, pp. 1 – 376, Chicago, 2007.

[8] Quan Vu, "VOS: The Corpus-based Vietnamese Text-to-speech System," Journal on Information, Technologies, anh Communications, 2010.

[9] Quan Vu et al., (2012). "Nghiên cứu xây dựng hệ thống Voice Server và ứng dụng cho các dịch vụ trả lời tự động qua điện thoại". Technical report, Research project, HCM City Department of Science and Technology, Viet Nam.

[10] Tran, T.K., Phan, T.T.: An upgrading SentiVoice-a system for querying hotel service reviews via phone. 2015 International Conference on Asian Language Processing (IALP). Pp.115-118.

[11] Richard Montague, Formal Philosophy: Selected Papers of Richard Montague. Bell & Howell Information & Lea, pp. 1 – 119, New Haven, 1974.

[12] Sandiway Fong, "LING 364: Introduction to Formal Semantics. www.dingo.sbs.arizona.edu/~sandiway ", 2012.

[13] Steve Young et al, The HTK Book (version 3.4). Available: htk.eng.cam.ac.uk/docs/docs.shtml, 2006.

[14] Thang Vu, Mai Luong, "The Development of Vietnamese Corpora Toward Speech Translation System," RIVF- VLSP 2012, Ho Chi Minh City, Viet Nam, 2012.

[15] Thien Khai Tran, Dang Tuan Nguyen (2013). "Semantic Processing Mechanism for Listening and Comprehension in VNSCalendar System". International Journal on Natural Language Computing (IJNLC) Vol. 2, No.2, April 2013.

[16] TK Tran, TCK Tran, TA Mai, NMH Nguyen, HT Vu. EDUVoice - a system for querying academic information via PSTN. The Third Asian Conference on Information Systems (ACIS 2014). Nha Trang.

[17] Tran, T.K., Phan, T.T.: An upgrading SentiVoice-a system for querying hotel service reviews via phone. Asian Language Processing (IALP), 2015 International Conference on, 115-118.

[18] TK Tran, DM Pham, B Van Huynh. Towards Building an Intelligent Call Routing System International Journal of Advanced Computer Science and Applications 7 (1).

[19] Tran, T.K., Phan, T.T.: A hybrid approach for building a Vietnamese sentiment dictionary. J. Intell. Fuzzy Syst. 35(1), 967–978 (2018).

[20] Tran, T.K., Phan, T.T.: Mining opinion targets and opinion words from online reviews. Int. J. Inf. Technol. 9(3), 239–249 (2017).