Voice of Southern Frontend Report

1^{st} Dang-Tam T. Le
Advance Program in Computer Science
Faculty of Information Technology
VNU-HCM, University of Science
ltdtam@apcs.vn

2^{nd} Do Tri Nhan
Advance Program in Computer Science
Faculty of Information Technology
VNU-HCM, University of Science
dtnhan@apcs.vn

***Abstract*—Many Text-To-Speech (TTS) models for Vietnamese language are developed and deployed by technology companies as well as Vietnam universities. This report aim to analyze the performance of VNU-HCM University of Science 's TTS system Voice of Southern compared with others in Vietnam. The authors proposed tests case which cover most of special and difficult Vietnamese sentences to conduct experiments on TTS systems.**

I. Introduction

A. Voice of Southern and other TTS systems in Vietnam

1) Voice of Southern System: Voice of Southern (VOS) is a TTS system developed by AILab from VNU-HCM University of Science. The first version of VOS (1.0) was released in 2019, followed by the 2.0 and 3.0 version in 2010 and 2012 respectively. The latest update of VOS is in 2017 with performance and quality improvement. The current VOS support instant response ($\frac{1}{15}$ realtime) with three different Vietnamese accents: Southern, Central and Northern. [1]

2) Other TTS systems: We divide list of TTS systems into two categories: one that supports Vietnamese and counterparts.

TTS system which does not support Vietnamese:

- Vocalware: Vocalware is a cloud based API for integrating real time Text-To-Speech into online web sites and mobile applications. Over 100 TTS voices in over 20 languages, offer APIs for multiple platforms, with demo Oddcast [2]
- Nuance TTS with 119 voices, 53 languages available to support, 25 years of Nuance TTS expertise [3]
- NaturalReader is a professional text to speech program that converts any written text into spoken words. It is a downloadable text-to-speech software for personal use, can read any text such as Microsoft Word files, webpages, PDF files, and E-mails, 74 TTS voices. [4]
- Microsoft Sam TTS Generator is an online interface for part of Microsoft Speech API 4.0 which was released in 1998. [5]
- Waston - IBM cloud: understands text and natural language to generate synthesized audio output complete with appropriate cadence and intonation, available in 13 voices across 7 languages [6]
- iSpeech : 27 languages, 3 reading speeds [7]
- Amazon Polly, uses advanced deep learning technologies to synthesize speech that sounds like a human voice, with 29 languages [8]

TTS system which supports Vietnamese:

- Text-to-Speech on Google Cloud, it converts text into human-like speech in more than 100 voices across 20+ languages and variants. It applies groundbreaking research in speech synthesis (WaveNet) and Google's powerful neural networks to deliver high-fidelity audio, with Vietnamese, there are 4 types of voice. [9]
- Responsivevoice is one of the most popular html5 text-to-speech API, support 51 languages. [10]
- VBee is a solution development, data digitalization and artificial intelligence leaded by Dr. Nguyen Thi Thu Trang. VBee covered 7 Vietnames accents. [11]
- FPT.AI Speech: Based on Amazon's cloud, supporting Nothern accent(male and female voices), Central (only male) and Southern (only female). [12]
- Text2speech of VTCC developed by Viettel which supports Northern accent (2 female voices and 1 male voice), Central voices are not supported and Southern (1 male and female voices). [13]
- eSpeak is a compact open source software speech synthesizer based in Vie-HTS project (Vietnamese Human-based Text-to-Speech) [14]
- Text2Voice developed by FIBO Tecchnology Company which allow voice's speed and automatic innotation adding [15]
- Vnspeak TTS by Le Anh Tuan - Vietnamese Multi-platform Text-to-Speech Engine, supports multiple platforms, such as Windows 32 bit, Windows 64 bit, Linux and other embedded systems. [16]

II. Experiements

We conducted experiments with the TTS systems which had proved to have significant result on basic TTS tasks. Furthermore, those systems are still updated and maintained. We choose 5 TTS systems:

TTS system that support Vietnamese

- Google TTS WaveNet [9]
- Responsivevoice [10]

TTS system that developed only for Vietnamese

- VOS [1]
- VBee [11]
- FPT.AI [12]

	Number of test case	VOS	VBEE	FPT.AI	Google	Responsive
Normal text	5	100%	100%	100%	100%	100%
Time-Date	15	73.30%	33,3%	53.30%	80% (it can specify the session of day base on the hour)	80%
Acronyms	12	41.6% (Some word is not spelled)	66.60%	83.30%	75%	75%
Proper Noun	5	20%	60%	40%	100%	100%
Address	5	20% (Miss some characters)	40% (spell each character)	40%	40%	40%
Unit of measurement	10	70%	100%	60.00%	70%	70%
Teen code	5	20%	20%	40%	40%	20%
Upper-Lowercase recognizing	2	50%	0%	50%	100%	100%
Mathematics	6	50%	66.60%	83.30%	83.30%	83.30%
Cross language	5	20% (Can not detect English words, read as Vietnamese)	80%	60%	100%	100%
Special case	5	60%	40%	60%	60%	60%
Total	**75**	**52%**	**59%**	**61%**	**73%**	**72%**

Table I
OVERALL RESULT

A. *Test cases*

The test cases we provide aim to test the performance of the VOS systems. It consist of many sentences which are used on regular basis. We categorize them into 9 types:

1) Normal text : Normal sentences which contains only letters.
2) Addresses: Addresses in Vietnam. This is a challenging task because of the inconsistent in the way to say an address.
3) Unit of measurement: This is a basic requirement of an TTS system because unit of measurement is widely used on variety kinds of document.
4) Time-Date: A basic and difficult requirement for TTS system. Two challenging problems that TTS system need to handle are the inconsistent in the way to read date, time and how to detect dates, time from the context of the document.
5) Acronyms: The main challenge with acronyms is whether the TTS system should pronounce the full word that acronyms stand for or pronounce letter by letter. We analyze the performance of those systems on this test case based on the context of the document.
6) Teencode: Teencode is words that used in social network and messenger applications, it contains acronyms combined with special characters. This is very difficult test cases, however, not a requirement for every TTS system to implement.
7) Upper-Lowercase recognizing: Some special words have different pronunciation when written in uppercase or lowercase, especially unit of measurements.
8) Proper Noun: Common English proper nouns.
9) Mathematics: Mathematics formulas.
10) Cross language: Document have some English words. These words is from the dictionary and name of person, place, etc,...

B. *Criteria*

- Pronunciation: TTS system need to read most of words in sentences correctly.
- Fluency: The system need to read sentences regular speed and pause when encountering commas and periods.
- Make the listener easier to understand, they don't try to guess the meaning of sentences leads to wrong understanding.
- Regional accents support: How many different accents each system support and the consistent in reading sentences by all supported accents
- Context: Result need to match the context of documents.

TTS system	Advantage	Foible
VOS	Good in special cases with many rules	Bad in containing English cases Bad in handle Abbreviations
VBEE	Quite similar to VOS, just different in rules for special case	Southern Voice is limited Bad in handling with numbers
FPT.AI	Smooth and fluently voice	Can't handle special cases
Google	Good at multi-lingual handling Handling the majority of cases	Unnatural
Responsive	Similar to Google Good at multi-lingual handling	Unnatural and not support many Vietnamese voices

Table II
ADVANTAGE AND FOIBLE

C. *Results*

1) Overall Results: Overall result could be seen in Table I

2) Detail Results: For the detail test cases and result, please see at: Test cases and detail results

3) Comparison: From test cases provided, we compare advantage and foible of each system in Table II

III. CONCLUSION

A. *Performance of VOS*

VOS system has natural and fluent voice. However, when dealing with informal document, VOS has few minor errors result in unnatural results. This could be explained by the rapid change in the informal language, especially ones used in Internet.

B. *Improvement Proposal*

From the reuslt of the experiment, we propose some improvment for VOS system:

- Using larger dictionary to cover more words and implementing module to recognize words which have different pronunciation in uppercase and lowercase.
- Update Acronyms and Teencode Dictionary
- Improve the processing special character module to produce the result to match the context of document.
- Improve the numeric processing module

References

[1] V. H. Quan. (2018, May) 20 years of vietnamese spoken language processing: Research achievement. [Online]. Available: https://ai4life.uet.vnu.edu.vn/wp-content/uploads/2018/05/AI4Life-Vu-Hai-Quan.pdf

[2] Vocalware. Vocalware's text-to-speech. [Online]. Available: https://www.vocalware.com/index/demo

[3] N. Communications. Nuance's text-to-speech. [Online]. Available: https://www.nuance.com

[4] N. Ltd. Natural 's text-to-speech. [Online]. Available: https://www.naturalreaders.com/online/

[5] Microsoft. Online microsoft sam tts generator. [Online]. Available: https://tetyys.com/SAPI4/

[6] IBM. Waston text to speech. [Online]. Available: https://text-to-speech-demo.ng.bluemix.net

[7] iSpeech. Demo. [Online]. Available: https://www.ispeech.org

[8] A. W. Services. Amazon polly. [Online]. Available: https://aws.amazon.com/polly/

[9] Google. Demo. [Online]. Available: https://cloud.google.com/text-to-speech

[10] Responsivevoice. Responsivevoice's text-to-speech. [Online]. Available: https://responsivevoice.org/

[11] VBee. Vbee 's text-to-speech. [Online]. Available: https://vbee.vn/

[12] F. CORPORATION. Fpt.ai speech. [Online]. Available: https://fpt.ai/tts/

[13] Viettel. Viettel 's text2speech. [Online]. Available: https://vtcc.ai/tts

[14] F. S. Foundation. Vietnamese human-based text-to-speech. [Online]. Available: http://espeak.sourceforge.net/

[15] F. T. Company. Fibo text-to-speech. [Online]. Available: https://fibo.vn/voice/demo/

[16] L. T. Anh. Vnspeak tts. [Online]. Available: http://www.vnspeak.com/vnspeak-tts/