Voice Mirror: Language Recognition Based Solution

Mansi Agarwal[1], Prof. Babar Ali[2], Prof. Abhishek Saxena[3]

[1]*Software Engineering,* [123]*Department of Computer Science and Engineering, Noida International University, Greater Noida, India*

Abstract-Allowing people to converse with machines is a long-standing dream of human-computer interaction. Have you ever talked to a machine? What if computer system start talking in human like language. Your own virtual personal assistant talking in familiar voice or your virtual assistant reading books/stories in your favorite person's voice. All of these things will be possible in today's era with voice mirroring. This paper proposes the idea and discusses the comprehensive insight into benefits of using voice mirroring with the help of speech recognition and synthesis and various futuristic applications to benefit the mankind.

Keyword-NLP, Speech Recognition, ML, Personal Assistant, Text-To-Speech (TTS)

I. INTRODUCTION

In today's world computers have become integral part of our lives. We use computers/ smart phones in day today life extensively to ease our work and to make it more efficient and fast. Making human computer interactions more human-human like is the most challenging problem of our current times and what better can be using human's voice to do the same. As it is more natural and human like, unlike typing instructions on console or filling up forms to do the same. But Processing human voice and converting it to a format computer understands itself is challenging task.

This human-computer interaction enables real-world applications like automatic text summarization, sentiment analysis, topic extraction, named entity recognition, parts-of-speech tagging, relationship extraction, stemming, and more.

II. SYSTEM MODEL

For a given System of humans A and B if network is trained for Human A's voice, and if any other human say B give input to nlp trained network it should spell out the given text input in Human A's voice.

Step 1: Decide the input language for network i.e. English

Step 2: Take input any user A voice with given alphabet.

Step 3: Train the network for user's voice with respective alphabet/ input space.

Step 4: Take Input from User B in text form.

Step 5: Use the trained model to spell out the given text in User A's Voice.

III. RELATED WORK

Ability of computers to understand natural speech has evolved a lot in last decade by using deep neural network i.e. Google Voice Search. However Generating Text To Speech (TTS) can be largely categorized in 3 categories.

1. Concatenative TTS

In this approach large DB of short voice fragments are recorded and then combined of utterance. However it makes it difficult to modify the DB as it needs re-recording of all the DB.

2. Parametrized TTS

Here instead of storing plain audio fragments, information is stored in data model which can be modified based on certain set of input parameters. However it sounds less natural than concatenative TTS.

3. Neural Network Approach

Extracting features from audio samples and using deep generating model to create raw audio, i.e. wavenet6 which gives more natural sound and easy to extend.

IV. PROPOSED METHODOLOGY

Step 1: Decide the input language for network i.e. English.

- First to train the model for English language then later on extend it for other languages.

Step 2: Take input any user A voice with given alphabet.

- Device the given voice input into lexicons and then form the parse tree or some other representation later on to be used for text to speech synthesis.

Step 3: Train the network for user's voice with respective alphabet/ input space.

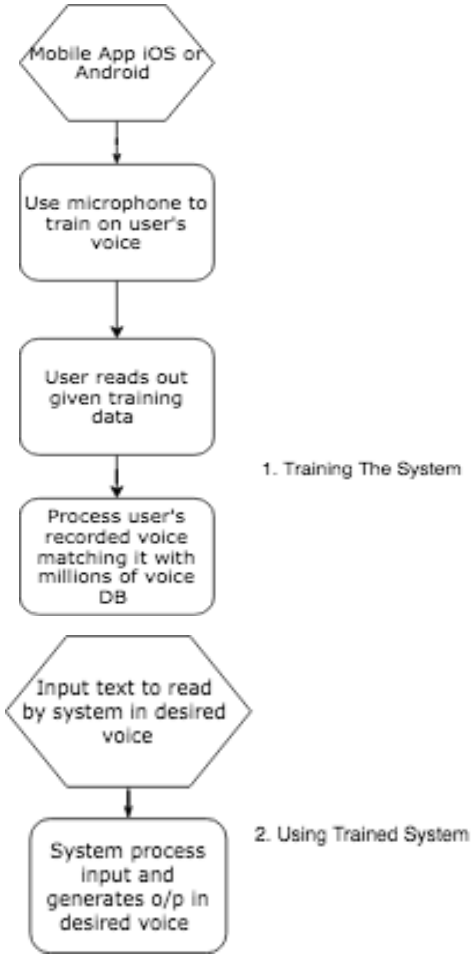
use the above intermediate representation for training purposes.

Step 4: Take Input from User B in text form.

Taking input from user in form of text from some defined interface may it be app or console or web page.

Step 5: Use the trained model to spell out the given text in User A's Voice.

Generate the output using the above trained model.



V. POSSIBLE APPLICATIONS

Possibilities with voice mirror are endless, below are some of them.

1. Mobile (iOS or Android) app to create someone's voice avatar .

2. Train on some singer's voice and get your own lyrics or songs sung by him with tap of a button.

3. Digital Voice Preservation

a. Any user can get his voice converted to digital mirror and get it stored on cloud, which can be used by millions of users for further usage.

4. Personalized Text Reader Extension

a. To be used as personal assistant for bedtime stories for kids who want to listen story in his mom's or dad's voice.

b. App extension for IOS or Android apps can be created, to be used with pdf or word documents in your own choice of voice.

5. Marketplace of Voice Mirrors

a. A curated catalogue of digitally processed voices, engineered and refined set of voices from independent and commercial solutions available to be bought by users worldwide.

VI. CONCLUSION

As Voice mirroring have various application including customizing your personal virtual assistant. This helps in creating more intimate user experience, which is missing in the present. This can not only improve the experience, it can revolutionize various industries. Voice mirroring will really take the personalization with machine and computers to a whole new level.

FUTURE SCOPES

Possibilities with systems with voice as input for interaction with computers are endless. The very need to of having the most fluid and easy to use medium of communication with machines makes it challenging yet super exciting piece of puzzle in computer science domain.

REFERENCES

[1] Gelbukh, "Natural Language Processing", Fifth International conference on Hybrid systems, Nov 2005
[2] Lawrence R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE, 1989
[3] Parsing: Dan Klein and Christopher D. Manning. Accurate unlexicalized parsing. In ACL, pages 423–430, 2003
[4] Kevin Knight and Daniel Marcu: Summarization beyond sentence extraction. Artificial Intelligence 139, 2002.
[5] Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, Koray Kavukcuoglu, "WaveNet: A Generative Model for Raw Audio"
[6] Deepmind Wavenet, "A Generative model for raw Audio"