International Journal of Engineering and Techniques - Volume 6 Issue 1, February 2020

Comparative Study on Adaptive Questionnaire Model Building and Mixed Reality Simulation for Recruitment Process

Amreen Hasheem Siraj, AneeshSanganeria, Avishek Mukherjee, Dr. Preeti Satish, Dr. Ramesh Babu

DAYANANDA SAGAR COLLEGE OF ENGINEERING

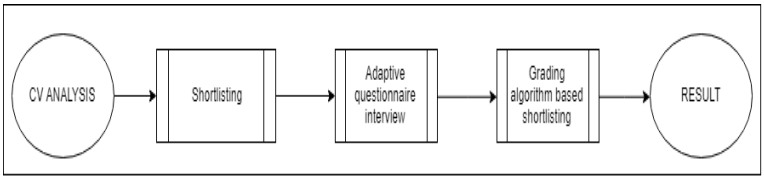
(2019)

Abstract – Artificial Intelligence hasn't left a facet of life untouched. As it makes its mark in every step of an organization, its application in the field of Human Resource(HR) needs critical analysis. Recruitment is the backbone of a well-functioning organization. In this paper, we will be highlighting the various methods to perform applicant resume screening, automating the technical HR round, engaging, shortlisting and hiring. Automating this process allows an organization to focus on building and regulating policies. This paper highlights the cascading impact of the various AI frameworks used in the HR recruitment process giving a deeper insight into it.Mixed Reality approach is used to simulate the virtual environment, giving a realistic feel to the applicant.

Keywords— Natural Language Processing, Artificial Intelligence, Neural Networks, Mixed Reality, Mixed Reality Architecture, Resume Screening, Automating Interview ,Virtual Reality, Immersive Environment

1. Introduction

As Artificial Intelligence and Mixed Reality is taking over every facet of life, it has left no stone unturned even in the field of Human Resources(HR). Our aim is to employ Machine Learning to perform resume screening, filter applicants with the required skills, automate the technical HR interview and recruit the most skilled applicants. Further examination of the paper provides a deeper insight on the various approaches that can be adopted to implement the various functionalities. Initially, screening of the applicant's resume is conducted by applying the Tesseract Optical Character Recognition(OCR) library which inputs an image and performs a grayscale function during preprocessing. The text is extracted and applicants with the necessary skills are filtered according to the company's requirement. This step is followed by an adaptive questionnaire segment that employs Machine Learning for the segment that throws questions under the same domain



as the applicant's previous answer. The interface provided to the applicant is by using Mixed Reality which simulates an interview environment in a VR headset with a virtual interviewer to simulate the personal interview experience. The different approaches used for conversion of text to speech are Unit Selection, Hidden Markov Model(HMM) synthesis and TD-PSOLA, all of which specify text-to-speech and speech-to-text conversion methods. However, the parameters for each

Fig 1. Smart Interviewer Application Flow

implementation change and the one with the best results has been selected. Further, we explain the different methods for model creation with increasing accuracy in keyword extraction which is a critical step in asking the next question. The final method chosen goes beyond keyword matching alone and chooses the next questions based on its proximity to the previous answer in terms of semantics. This entire experience will be simulated in a VR headset using Mixed Technology. The engine used to build the application is Unity 3D along with MR plugins. A comparative study of the various types of MR implementations such as Enhanced, Immersive and Blended gives a deep insight on the best implementation method to achieve a simulated virtual

environment. This survey paper gives a detailed account of the various approaches that can be used to implement the different stages in Fig 1.

2. Index

3. Literature Survey

3.1. Natural Language Processing

Natural language processing is a field which involves computational processing. It consists of a variety of topics which involve understanding human languages. It relies on computational methods which are data driven involving probability, statistics and machine learning. Since there have been many increases in computational power which are controlled by Graphical Processing Units, it allows the use of deep learning which can utilize artificial neural networks. This lets machine learning technique use millions of trainable parameters. Since there are a large number of datasets which are accompanied by a very detailed data collection processes, we can use them to train deep learning data algorithms. This has led to improvements both in areas which directly applies to achieve practical and useful objects as well as in core areas of NLP.

3.1.1. Text-to-speech(TTS)

Text-to-speech's goal is to create natural sounding speech from text. The advancements in Text-To-Speech has led it to produce speech in different speaking styles with different speaker characteristics and emotions. Speech synthesis is the process of generating human voice artificially either in the form of speech or text. The computer system used for this process is called as speech synthesizer. The TTS system converts normal text to human speech. The goal being to automatically produce speech from any arbitrary sentences. The TTS procedure contains of two main phases:

1) Text analysis: Input is transcribed into a phonetic representation.
2) The generation of speech waveforms, where the output is produced from the information obtained from the first step.

Simplified version of synthesis procedure is presented in Fig 2.

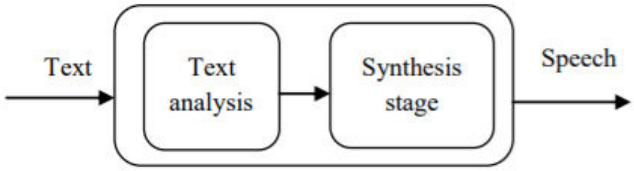


Fig 2. Phases of Text- to- speech system[6]

The quality of a speech synthesizer is measured based on two primary factors:

1)Similarity to normal human speech.

2)Intelligibility(ease of understanding)

Typical text to speech has a front end and a back end. Front end is responsible for text normalization and text to phoneme conversion. Text normalization is also called as tokenization and this is responsible for the numbers and abbreviations in the raw text to be converted to written words. Text to phoneme conversion is the process of assigning phonetic transcriptions to words and dividing them into phrases. The output of the front end is a symbolic linguistic representation.

Development of speech synthesis is complicated and consists of the following challenges:

1. Developing a text to speech system requires knowledge of how speech is produced and about how languages are developed.
2. Good software skills are required to implement a fully functional system.
3. Qualities of speech synthesis include naturalness and intelligibility. Naturalness can be defined by how close the output of the sounds are related to each other. Intelligibility

can be defined as the ease with which outputs are understood.

3.1.2. Methods of text to speech synthesis

i) Formant Synthesis

Formant synthesis is a method that consists of a set of rules that are used to control a highly simplified source-filter model. The model is based on the possibility of treating the source(glottal) as completely independent from the filter. The parameters are considered to be frequency and bandwidth and the operation of the filter is based on this.The resonance on the vocal tract influences the values of the formant frequencies.

Formant synthesis method does have the advantages of working with modern computational resources, but it cannot produce natural sounding speech.

ii) Articulatory Speech Synthesis

This type consists of biomechanical models to be used in the production of speech. The effect of organs that produce speech for example tongues, lips can be simulated by these models.

Execution of the models need output of models to not be measured similar to natural sounding fluent speech.

iii)Concatenative speech Synthesis

The concatenative method uses phonemes which are recorded snippets of smallest speech units. Phonemes consist of vowels or constants. Phonemes need high quality equipment to record in an area which is sound-proof to develop a voice database. It is stored in the form of a waveform.When the synthesizer receives this text, it will first map all the characters to its phonemes and then assembles the respective phonemes to generate the speech. Recordings of actual sounds are used as the speech is generated to make it sound as natural as possible.

3.1.3 Text to speech based on the use of syllables

Text to speech(TTS) systems are categorized into rule-based synthesis and concatenative synthesis. A rule-based synthesizer consists of a set of rules that define an explicit knowledge of what mechanisms can be used for the production of speech. Concatenative speech synthesis approach dictates that “no” is a necessity to establish the speech production rules. Hence concatenative synthesis is considered to be easier that rule-based synthesis. Concatenative synthesis combines prerecorded synthesis units such as words and syllables.

3.1.4. HMM (Hidden Markov Model)

Hidden markov models are formed using parameters which consist of speech data. The major advantage of HMM-based speech synthesis is that their size of footprints are significantly lower, so that these systems can be used in hand held devices.

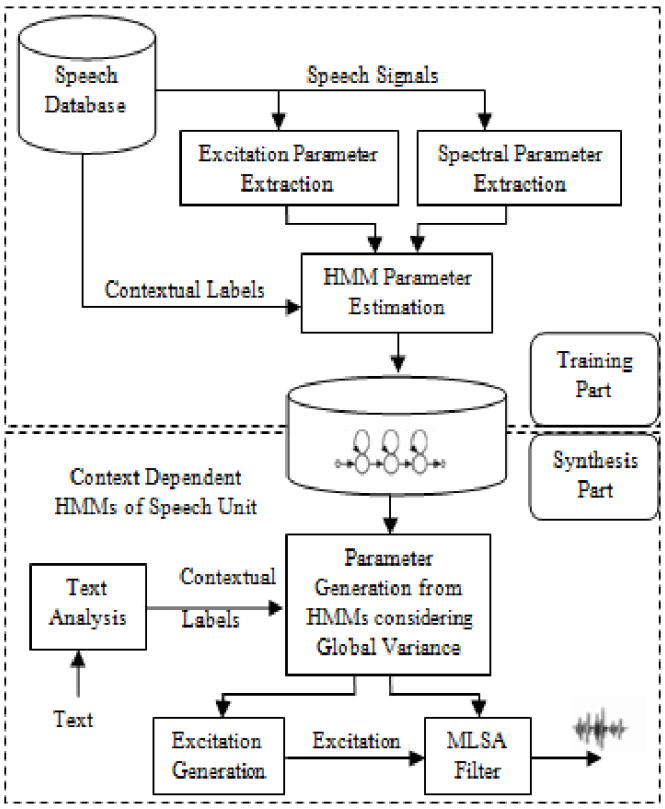


Fig 3. HMM based speech synthesis model

3.1.5. TD-PSOLA model

TD-PSOLA is an approach which can be used modify scale of pitch. The duration can be scaled without the loss of any source or information of system. There will not be any significant distortion in the synthesized speech since the modification is carried out at the signal level. TD-PSOLA is used to synthesize and make changes in outline of the pitch in the neutral speech.

3.1.6. COMPARISON OF TECHNIQUES

To compare the techniques, we chose effectiveness, flexibility and simplicity to be the performance parameters they will be graded on. Under effectiveness performance factor we will be focusing on the naturalness of the speech which would include how human-like the voice would sound and how less noise it carried with it. Flexibility is measured on the basis of difficulty to edit the system when applied to an application and adjust its parameters. Simplicity is measured on how well it can understand what we represent on the complexity of an implementation.

Techniques	Performance		
	Effectiveness	Flexibility	Simplicity
Unit Selection	Speech quality has better naturalness than the conventional techniques.	-	Difficult to produce voice quality variations
HMM	Synthesize highly intelligible and smooth sounds	Changes its voice characteristics and emotions	Relative ease with which HMM based systems adapted to speakers not present in dataset
TD-PSOLA	Gives good result when used on both speech signal		Ease of Implementation

Fig 4. Performance Comparison Table[5]

The Hidden Markov model synthesis is the technique suitable for our project as it allows more variations in the characteristics of voice and it can portray emotions. That is vital in making our simulated assistant seem much more realistic as the user must get as close to an interview experience as possible. TD-PSOLA gives a good voice quality but the HMM (Hidden Markov model synthesis) has most flexibility to change its voice characteristics. Unit selection is limited in this field as it cannot vary its voice patterns by a significant amount.

3.2 Related Keyword Extraction From image

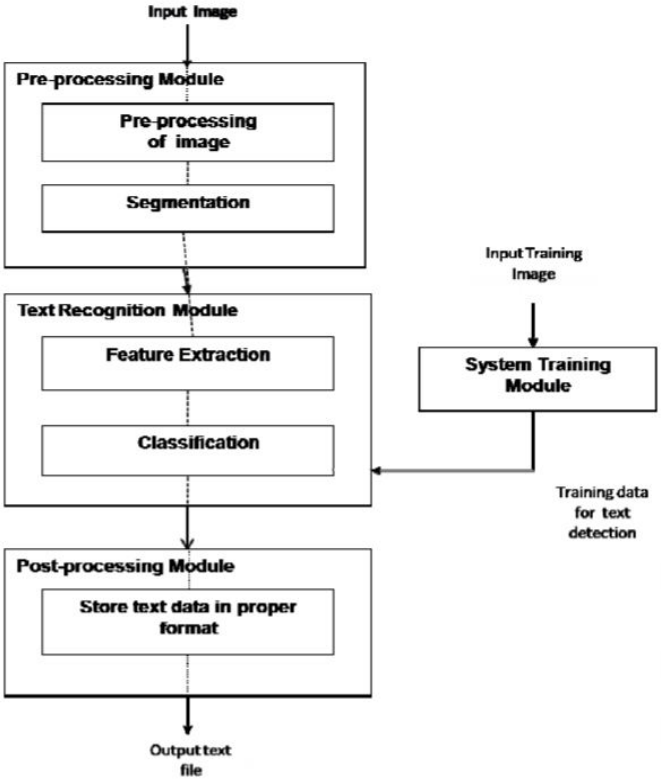


Fig 5. Architecture of image to text[10]

The image to text conversion takes place in 4 steps:

- Pre-Processing module
- System training module
- Text recognition module
- Post-Processing module

3.2.1 Pre-Processing module

The resume is largely filtered by the optical scanner and is changed into the type of an image. An image is the combination of picture components which are called pixels. The pixels contain fundamentally two qualities ON and OFF. The ON worth focuses that the pixel is active and the OFF worth focuses that the pixel isn't active. At this stage we have the information as an image and this picture can be additionally investigated with the goal, i.e the significant data can be recovered, in order to improve the nature of the picture and make it appropriate for further examination, we play out some activity on it, for example; Grayscale change, Binary

picture transformation and Classification[10](of which classification is the most significant).

3.2.2. System training

This module can be used to train the system for text recognition. Before converting the printed documents into editable and searchable documents, the first and mandatory step is providing training to the system. Here, during training the font followed in the scanned document should be identified by the interviewer. Then the interviewer types all the characters that are required for recognition from the scanned document as an image file. This image file should be provided as an input during the training process.

3.2.3. Text recognition module

This module can be utilized for content acknowledgment to get a pre-prepared image and provide information which is reasonable in structure. Subsequently, in this module the following strategies are utilized:

i) Feature Extraction

Highlight extraction in the procedure to recover the most significant information from the crude information which implies that it is based on the characters which can be spoken precisely. To store the various highlights of a character, different classes are made. There are numerous strategy utilized for highlight extraction like Principle Component Analysis (PCA), Linear Discriminant Analysis (LDA), Independent Component Analysis (ICA), Chain Code (CC), Zoning, Gradient Based Highlights, Histogram, etc.

ii) Classification

Classification is a method to recognize each character and appoint it to the right character class, so that messages in pictures are changed over in to PC justifiable structure. This procedure utilizes the removed components of constituent images in order. For example; contribution to this stage is yield of the element extraction process. Classifiers contrast the information, include and exclude examples, hence, discovering the best coordinating classes for input. There are numerous methods utilized for arrangement, for example, Artificial Neural Network (ANN), Template Matching, Support Vector Coordinating (SVM), etc.

3.2.4 Post-Processing Module

The field of Text Recognition Module is in the structure content information which is comprehended by the PC. So, it is essential to store it into an appropriate software (for example txt or MS-Word) for more distant utilization. For example, Editing or Searching in that information.

The Paper record is thoroughly examined by the optical scanner and is changed into a type of an image. An image is a combination of picture components which are called pixels. At this stage we have the information as a picture and this picture can be additionally broken down with the goal to recover significant data.

3.3. Machine Learning Models

Machine Learning concepts are used to extract skills from the applicant resume, provide questions by the interviewer and document responses. The technique wherein an interviewer asks questions based on the interviewee's response is emulated by the model. Keyword extraction forms an essential step in the model implementation as it extracts keywords from the interviewee's response to provide further questions. Furthermore, we focus on the semantics of a phrase rather than just matching keywords.

Machine Learning employs supervised and unsupervised learning, however, keyword extraction prefers supervised learning. Supervised Machine Learning techniques include a model that is trained on a set of keywords. It requires the author to manually specify the keywords in a document which is extremely tedious. Training of a model can be extremely domain-specific and the model requires to be retrained every time the domain changes. The following approaches iterate the different methodologies to implement the above techniques.

3.3.1. Statistical Models

This approach is simple and does not use any training dataset. This method is language and domain-independent. It is a very popular method to employ statistics to words in a document to identify keywords. The most widely used method sis term frequency-inverse document frequency (TFIDF) which is a numerical statistic that reflects the level of importance of a word in a document. It increases by the number of times a word repeats in a document. Similarly, statistical methods like: n-gram statistics, Patricia Tree (PAT tree), word co-occurrences and word frequency have been

used. The disadvantage lies in the fact that technical words, like regression, which are very important may occur only once in the entire document. Usage of statistical models may inevitably filter out these words. [1]

3.3.2. Conditional Random Fields:

Conditional random field provides sentence segmentation and labelling. It showed better results than most statistical approaches applied to keyword extraction. The main steps in this model are (i) Feature extraction (ii) Pre-processing (iii) Model training (iv) Keyword extraction and (v) CRF labelling and (vi) Results evaluation. [3] CRF++ tool used Part of Speech tag to extract keywords from documents. [4]

3.3.3. Support Vector Machine(SVM), Multiple Linear Regression(MLR) and Logistic Regression(Logit)

SVM, MLR and Logit has been used for keyword extraction with TFIDF as a baseline for the process [3]. SVM is a supervised learning technique that categorizes a given set of inputs and produces optimal outputs. SVM approaches keyword extraction as a classification problem, wherein the words in a document can be classified as: 'good keyword', 'indifferent keyword', and 'bad keyword'[5]. We give a specification of keyword in this paper. In this approach, candidate keywords are selected by tri-grams and then define the features as 'global context information' and 'local context information'. The keywords are extraction by a classification model that is trained in advance. Results show that SVM show better results than CRF and logit shows better results than MLR. [5]

3.3.4. Rapid Automatic Keyword Extraction algorithm (RAKE)

Rake is a domain independent keyword extraction algorithm which tries to determine key phrases in a body of text by analyzing the frequency of word appearance and its co-occurrence with other words in the text document. It operates on individual documents and enables application to dynamic collections. It operates well on multiple types of documents and can be easily applied to new documents. RAKE is based on the assumption that a keywords contain multiple words but very few stop words as they are considered meaningless. RAKE uses phrase delimiters and stop words to parse the document into candidate keywords. RAKE begins its extraction process by partitioning a document's text into a set of candidate keywords. First, the entire document is split into an array of keywords. The array is further split into sequences with stop words are used as delimiters. Words within a sequence are collectively considered a candidate keyword. A score is given to each keyword based on three parameters: (i) Word frequency (ii) word degree (iii) Ratio of degree to frequency. These parameters favor words that occur often and in longer candidate keys. The score for each candidate key is calculated as the sum of its constituent scores. After the scores have been given to the candidate keys, the top one-third scoring candidates are selected as keywords for the document.

3.3.5 Clustering Approach

The disadvantage of TFIDF and other statistical methods that it gave a score to keywords based on the frequency of occurrence without considering the words that are similar to it in terms of semantics. Moreover, for short documents TF may not be a good indicator for the importance of a keyword. This is an unsupervised approach which accounts for the frequency of other similar words. All the words are grouped into clusters in an unsupervised manner. If the total frequency value of a given cluster is relatively high, then it is likely that this cluster contributes the most to the current topic. Thus, the words in this cluster are given higher weights. The SRI Language Modeling(SRILM) toolkit is used for automatic word clustering over an entire document. [9]

3.3.6. Graph Based methods

For the purpose of keyword extraction, we use the iterative reinforcement approach to leverage sentence information. This algorithm is based on the fact that important words/phrases are connected to other important words/phrases in a graph. Four graphs can be created: (i) A sentence-sentence connected graph, (ii) A word-word connected graph, (iii) A sentence-word and (iv) A word-sentence graphs both having unidirectional edges.

(i) W-W Graph: There is no connection between words, the relation is given by a diagonal matrix. This is similar to giving an initial weight based on TFIDF and using an identity matrix for the word-to-word connection. (ii) S-W and W-S Graphs: The weight for an edge between a word and sentence is the TF of the word in

the sentence times the words Inverse Document Frequency (IDF) value.

(ii) S-S Graph: The sentence node uses a vector space composed of the weights of the S-W graph.

3.3.7 Word2Vec

One-hot representation is the most common representation of distributional semantics. Representation of this vector consists of 1's and 0's and dimensionality is equal to vocabulary's cardinality. Word embedding capture attributional similarities between vocabulary items. Similar words should be close to each other in the projected vector space i.e. grouping of vectors in the same space must share the same semantic properties.Latent Semantic Analysis (LSA) uses a counting base dimensionality reduction method in word embedding. Word2Vec offers low dimensionality and can help to reduce computational complexity. On comparison with distributional semantics methods, it causes less overfitting. Word2Vec is used to detect analogies between words as well. The model that we have built for our project represents a word in a vector space. We specify a threshold of count of words because words that are repeated once or twice in a large corpus are usual for the model and there isn't enough data to perform meaningful training on the data. After these specifications, the PageRank value of each word is calculated which works on random walk. The cosine distance is used to calculate the edge weights between the nodes which are the Word2Vec representations of a word. The words which have the highest PageRank are the keywords of document.

3.4. Mixed Reality

In mixed reality the capabilities of Augmented Reality(AR) and Virtual Reality(VR) are blended, bringing together physical and digital worlds to produce an environment where these objects coexist and interact in real time. Mixed reality is an important segment in our project where we are adding an immersive experience for the applicant. We transform the applicant's environment into an interview scenario where the applicant can experience the interview process. The experience entails a virtual interviewer that will ask questions and the applicant will answer those questions giving a real time feel to the applicant.

3.4.1. MR architecture

A significant quality of a well-structured architecture for MR composing devices is adaptability in adjusting and broadening the devices for new classes of creators[2].

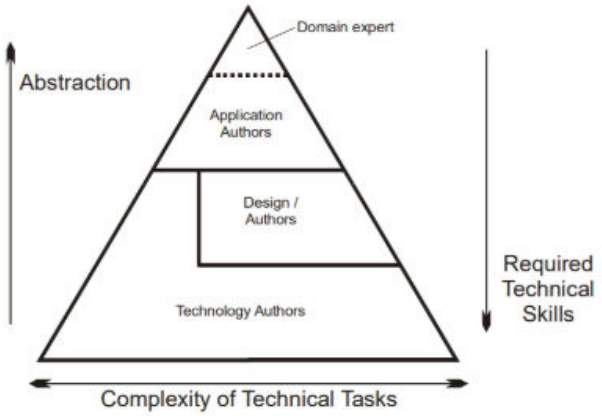


Fig. 6 Degrees of technical abstraction in MR[2]

Simply giving client explicit instruments doesn't ensure a simple and productive MR application advancement, since the approach to build the uses of this sort is unpredictable. There is no built up nor method that's understood on how MR applications ought to be assembled. Our procedure controls the creator bit by bit through an improved work process and guarantees that no stages required are excluded by the creator[2]. The principle thought in our methodology is to give the creator a chance to amass a MR application by the reuse (and adjustment) of predefined building squares (area vague MR segments that can be loaded up with area explicit substance). The granularity of these squares ended up being fundamental, since the creator must have the option to redesign the extent of his particular MR application into those MR segments.

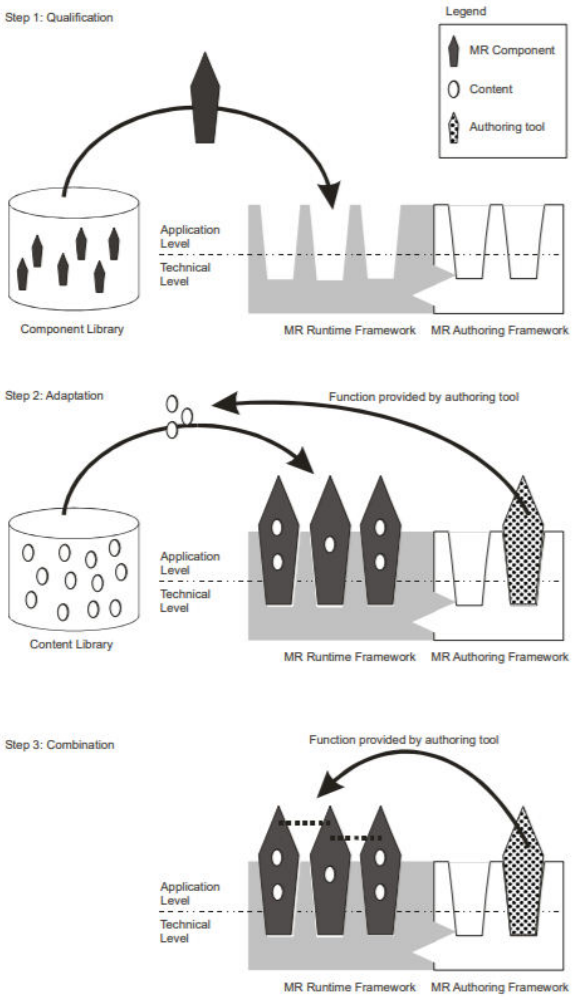


Fig. 7 Phases during MR development[2]

In our improvement procedure we distinguished four significant stages:

• Capability of MR segments

• Adjustment of the MR selected segments

• Blending of these MR parts to cooperate with one another

• Alignment of genuine and virtual items in MR

In the capability stage, the creator needs to recognize MR segments (ie. In this scenario an interview room, an interviewer and various other details) that are fitting with respect to the application goals. The MR segments are given by an authority gathering of MR objects. Taking into account the particular needs of various application areas, the convenience and relevance of those building squares must be assessed by those objects in the setting of exhaustive necessity buildings.

During the adjustment stage the creator has to modify the MR segments to fit the specific needs presented by other structure squares and by the application itself. Since the individual structure squares are generally made with various application settings in mind, which implies that they need to suit an assortment of needs. For instance, a MR part that comments on a genuine object (i.e. a table in the room or the interviewer) with 3D content could be modified by characterizing the shade of the content or the textual style and size. Application explicit content can be additionally made accessible as a part and may be put away in a focal part library. The substance parts can be utilized by the creator to alter MR parts for a particular application.

In the blending stage, the creator characterizes the connections and collaboration instruments between many MR segments. This assignment is actually bolstered because of the way all the MR parts are implanted in a typical MR run-time structure. All the MR segments have agreeing interfaces that take into consideration the correspondence with other MR parts. Our advancement procedures need to reflect and satisfy the addresses of the particular necessities of MR[2]. Since, in MR the arrangement of virtual items (content, video, explanation, 3D geometries) to reference objects in a genuine space is a vital factor.

We presented an extra fourth stage, the adjustment stage. The usefulness of apparatuses, that address this stage, plan to adjust the virtual items at the right position (for example, the creator indicates that a virtual interviewer may be adjusted to a seat).

3.4.2. Mixed Reality Methods -

- Enhanced environmental Applications
- Blended environmental Applications
- Immersive environmental Applications

Mixed reality is implemented using MRTK(Mixed reality toolkit) is an open source toolkit that has been around since the HoloLens was first released. The unity 3d engine provides support for building Mixed Reality.

3.4.3. Enhanced environment

One of the most powerful ways that mixed reality can bring value to users is by facilitating the placement of digital information or content in a user's current environment. This is an enhanced environment app. This approach is popular for apps where the contextual placement of digital content in the real world is paramount by keeping the user's real world environment "present" during their experience. This approach also allows users to easily move from real world tasks to digital tasks and back easily, lending even more credence to promise that the mixed reality apps which the user sees are truly a part of their environment. These types of apps can only be implemented and used using a Microsoft HoloLens.

Example use cases-

- Mixed reality cooking app which can be placed on top of your stove for cooking recipes and instructions.
- Mixed reality games to be placed on your tabletop for a real time board game experience

3.4.4. Blended environment

Given Windows Mixed Reality's ability to recognize and map the user's environment, it is capable of creating a digital layer that can be completely overlaid on the user's space. Thin layer respects the shape and boundaries of the user's environment, but the app may choose to transform certain elements best suited to immerse the user in the app. This is called a blended environment app. Unlike an enhanced environment app, blended environment apps may only care enough about the environment to best use its makeup for encouraging specific user behavior (like encouraging movement or exploration) or by replacing elements with changes (a kitchen counter is virtually skinned to show a different tile pattern). This type of experience may even transform an element into an entirely different object, but still retain the rough dimensions of the object as its base. Example use cases:

- Mixed reality application for interior designing.
- Mixed reality application for furniture placement and room decoration.

3.4.5. Immersive environment

Immersive environment apps are centered around an environment that completely changes the user's world and can place them in a different time and space. These environments can feel very real, creating immersive and thrilling experiences that are only limited by the app creator's imagination. Unlike blended environment apps, once Windows Mixed Reality identifies the user's space, an immersive environment app may totally disregard the user's current environment and replace it whole stock with one of its own. These experiences may also completely separate time and space, meaning a user could walk the streets of Rome in an immersive experience, while remaining relatively still in their real world space. Context of the real world environment may not be important to an immersive environment app. Example use case:

- Mixed reality application for tour of a place
- Mixed reality games to put yourself in a real time scenario.

TYPE	ENHANCED	BLENDED	IMMERSIVE
Overlay	full	mixed	full
Engine	Unity 3D, unreal engine, 3D max	Unity 3d,unreal engine	Unity 3D
Hardware	Only Microsoft holo lens	Hololens, vive, oculus	Hololens,vive ,oculus

Input possible	Gesture, Voice,eye tracking, spatial tracking	eye tracking, spatial tracking, Hardware	Gesture, Voice,eye tracking, spatial tracking input

4. CONCLUSION

From the insides drawn from the above survey analysis we can conclude that for text to speech the Hidden Markov Model Synthesis is the most flexible, simplest and the most effective among the three techniques. It has the ability to change its characteristics and speech patterns of its voice and is the technique we find to be the most compatible for our project. As we have interpreted that machine learning tools is a method that gives the best method for keyword extraction and preceding questionnaire, the methods such as keyword extraction or maximum word frequency as important keywords was not domain specific. To solve this we used Word2Vec model that understands the semantics of a phrase and allocates weighs to words based on semantic importance of words in that phase. These types of extraction based on allocating weights to important words are used in the adaptive question and answer section. Based on the comparison analysis we are using the immersive type of Mixed Reality(MR) as it precisely adapts to our requirement specifications.

5. REFERENCES

1. P. Chen, S. Lin, "Automatic keyword prediction using Google similarity distance", presented at Expert Syst. Appl., pp. 1928- 1938, 2010.
2. R.Dörner, M.Haller, “Efficient mixed reality application development” article ‘2004
3. A. Hulth, “Improved Automatic Keyword Extraction Given More Linguistic Knowledge”
4. H M Mahedi Hasan, “An empirical study of important keyword extraction techniques from documents”, 2017
5. Ramita P. Karpe, “A Survey :On Text To Speech Synthesis”
6. Desai Siddhi, “Survey on Various Methods of Text to Speech Synthesis”
7. KuoZhangHuiXuJieTangJuanzi Li, “Keyword Extraction Using Support Vector Machine”
8. Y.B. Wu; Q. Li; RSBot; X. Chen, “Domain-specific Keyphrase Extraction”, Information Systems Department New Jersey Institute of Technology Newark, NJ 0710
9. FeifanLiu,”Unsupervised Approaches for Automatic Keyword Extraction Using Meeting Transcripts “
10. P.M Manwatkar,S.Hyadav “Text Recognition from Images”,(ICIIECS)2015