“IntelliOrder” - Intelligent Speech Recognition Self-Ordering Management System for Restaurants

Muhammad Daniel Vineshen, Nur Khairunnishazainal and Minnu Helen Joseph

Abstract—This paper proposes a software system that simplifies the way orders are taken in a restaurant by developing a self-ordering system that is equipped with voice recognition that is able to capture the customer’s native language and translate it into the default language of the system. Moreover, the software system provides the capability of splitting bills for respective customers under the same table to save the hassle for the customers to split it among themselves. Furthermore, the software system will interconnect between multiple interfaces that will be used by six major users which are the customers, cooks, cashier, servers, managers/owners and administrator. Each of the users will be restricted to functionalities that are designated for them.

Keywords—Food Industry, Artificial Intelligence, Mobile Applications.

I. INTRODUCTION

Malaysia is a developing country that is adapting to a faster pace of lifestyle. Therefore, an implication of this new trend is Malaysians are eating out more frequently at various restaurants to avoid spending time preparing food as they are always busy. This has led to a great spike in business opportunities in the food industry sector [1]. The conventional way of restaurants in handling their day-to-day operations are outdated and inefficient as it is prone to countless errors and mistakes. The operations are done manually by the employees which are a waste of human resources as it can be enhanced and automated by using cutting-edge technology. Furthermore, increased globalisation has impacted growing businesses to adapt with the different linguistic boundaries as a mean of communication for managers and employees with their customers [2]. Thus, IntelliOrder is a software system that simplifies the way orders are taken in a restaurant by developing a self-ordering system that is equipped with cutting-edge functionalities.

II. LITERATURE REVIEW

This part covers numerous scholar journals related to the author’s project domain research which are mostly in the context of hospitality in the perspective of food industry such as restaurants.

2.1 Language Barrier or Cultural Differences

Language acts as a medium for a person to convey stories, socialize with friends, express themselves emotionally and share knowledge and expertise. Language symbolises someone’s self-identity that is cherished and valued by individuals all across the globe. Language acts as a bridge for people to understand each other better and
relate to their frame of mind [3]. The social identity theory discusses the physiological reasoning on why discrimination of intergroup occurs. The social identity theory states that people feel a greater sense of belonging when they are grouped with people they belonged to as it significantly increases their pride and self-esteem [4].

Social identity theory suggests that being unable to communicate or identify other cultural groups leads to negative outcomes. Social identity theory is based on the rationale that people are more attracted to people of similar characteristics and attributes which leads those individuals to feel more positivity when interacting with one another. People prefer to speak to individuals that are perceived to be more predictable and approachable. Thus, people prefer to encounter with individuals that they perceive are similar to them as it is more favourable and are evaluated highly [5].

However, studies that were conducted on intercultural groups contradicts with the social identity theory as consumers using intercultural services tends to be more linear with their service evaluation standards and are more tolerant and considerate in intercultural service encounters [4]. In contrast, a finding founded by researcher [6] that was conducted in the context of a waiter and a customer suggests otherwise, stating that intercultural services do influence the perception of customers when approached by services from a different culture. More so the customer’s response when encountering a service failure from a different culture [6]. However, all the previous studies stated did not take the customer’s emotional and cognitive mechanisms into consideration when interacting with intercultural services.

English as a Second Language (ESL) can be categorised as low literate consumers in regards to their English language level. According to Adkins & Ozanne, low literate consumers face many difficulties such as identifying the incorrect products needed and misinterpreting information about prices. Likewise, Viswanathan, et al concluded that low literate consumers are in a loss when making decisions on what to buy as their effort versus accuracy trade-offs are not formidable [7]. ESL customers face a lot of communication anxiety during intercultural services in a restaurant. ESL customers are more conscious of their lack of English skills and are self-restricted to certain actions due to that reason. ESL customers may find difficulties in clarifying their doubts as well as communicating in acquiring more information when ordering with the restaurant’s waiter or waitress. A qualitative study suggests that language barrier causes negative emotional and cognitive responses from customers as they restrict themselves when being confronted with intercultural services [8]. Hence, language plays a critical role in determining the satisfaction level of the customer’s experience during their stay in a restaurant.

2.2 Dimensions of Time

Dimensions of time in regards to pace and duration have been ventured and studies across various different domain that can be categories into different aspects of human life such as perspective of culture, communication and organizational. Aspects of dimensions of time include flexibility, linearity, pace (sometimes known as ‘rate’), scarcity, urgency, punctuality, scheduling, duration (sometimes known as ‘allocation’), synchronization and perspective of time [9]. The aspects that are relevant to this project are pace and duration. Pace can be defined as the rate it takes for a task to be accomplished. Conversely, the duration can be defined as the length of time taken in order for tasks or activities to complete, or the time allocated for a task or activity. Both of this aspects are
interconnected despite being their own separate entities [10]. However, the duration is reliant on pace. In the context of a restaurant, a slower pace in addressing the orders of a customer extends the customer duration at the restaurant. Hence, with the conclusion that a slower pace of service, resulting in more time wasted, which in return prolongs the duration of the encounter [10].

According to research conducted by researcher [11] indicated that consumer pays close attention to both pace and duration during a service encounter. There have been multiple kinds of researches conducted by different researchers that shows negative relationships between wait time and the customer satisfaction towards the service in multiple service contexts such as clinics, supermarket including fast food restaurants [11]. Naturally, the longer the duration of waiting time the perceived pace of the service the customers hopes to obtain will lead to customer feeling discontent. The previous statement is emphasised within the context of casual dining outlets as reasonable service times are crucial in order to compete with other establishments that have adopted fast-food models. As proposed by researcher [9] [10] that when service pace is slower than the expectation of the customer will lead the customers to feel frustrated and uneasy. On the contrary, when the service pace is faster than what the customer was anticipating will lead the customer to be unable to thoroughly enjoy their dining experience because of the feeling of being rushed [12].

2.3 Service Quality on Customer Satisfaction

Preceding studies have agreed that service quality and food quality plays a major role in the restaurant industry. Furthermore, the service quality and food quality of a restaurant greatly influences the customer’s satisfaction level and ensuring the retention of customer’s loyalty. Thus, these important qualities may lead to either positive or negative behaviour of the customers after their time at a restaurant depending on their satisfaction [13] [14]. Great quality services will lead to greater levels of customer satisfaction. Satisfied customers may revisit, spread positive news about the restaurant by mouth to others, meanwhile, dissatisfied customers will do the opposite and ruin the restaurant image [15]. Service quality is defined as the perception of customers on the excellence of the service provided if it meets the quality expectation of the customer [16]. Service quality is identified as the key factor in attaining customer satisfaction. Great service quality will ensure a competitive edge in the market by attracting customers and attaining high profits [17]. SERVQUAL scale was developed in the conquest of reliably measuring service quality in service-related situations. SERVQUAL is made up of five components which are reliability, responsiveness, empathy, assurance and tangible. As stated, one of the factors is responsiveness. The component that is relevant to this project is responsiveness. Responsiveness is classified as the eagerness and timely services provided to customers. Therefore, providing customer services in a restaurant in a timely manner is critical in increasing customer satisfaction [17].

Customer satisfaction can be explained by using the Expectancy-Disconfirmation theory states that customers measure their satisfaction via real-life experience or performances and compare it with their expectations [13]. This theory states three possible evaluation outcomes by customers: (1) neutral confirmation when performance is as expected; (2) positive confirmation when performance is better than expected which in return lead to satisfaction; (3) negative confirmation when performance is worse than expected which in return leads to dissatisfaction [13]. As
a result, every organisation needs to ensure their service quality standards are on par with the expectation of the customers in order to satisfy the customers need and retaining their loyalty.

2.4 Service Encounter Concerning Customer Satisfaction

The customers of a restaurant usually begin evaluating their dining experience at a restaurant at any given time in regards to the aspects of service scape and the services provided by the staff of the restaurant [18]. Service scape is a model proposed by Booms and Bitner regarding the correlation between the services interaction between a customer and the service provider against the physical environment the service interaction is occurring [8]. Service encounter also referred to as “moment of truth” is defined as an interaction between a customer and a service provider that can be of either inanimate objects such as the service scape or with the staffs of a service organisation [18]. Service encounter states that a customer will begin evaluating the services provided by the service provider at any point of time, thus, emphasizing the importance of service quality during a service encounter as well as services cape encounter [9].

According to researcher [19], there are three phases that undergoes during a service encounter: (1) early phase where the commitment of the service encounter occurs such as the customer enters the restaurant and awaits their order to be taken, (2) middle phase where the main purpose of the service encounter is completed such as the customers have made their order and are consuming the food and drinks ordered, and (3) last phase that concludes the service encounter such as paying at the cashier of a restaurant. In relation to the previously discussed findings about the duration dimension of time regarding the interacting outcome of service stage and service delays on consumers’ reactions to waiting can be perceived that customers in the context during a dining experience at a restaurant finds disappointment during the early and middle phase of the service encounter despite the duration of the delays were same for both phases [19].

According to a research conducted by researcher [20] that identified that communication is the first most important factor that customers anticipate during a service encounter to enhance customer relationship. Customers expect to communicate in an understandable way, able to acquire information that clearly articulated, and is shown appreciation by the staffs of a restaurant. The same research indicates that trust is the second most important factor. Customers hopes for the services provided by the restaurant to be trustable in order to enhance customer’s confidence in that particular restaurant. In order to gain the trust of the customers, transparency and obligations must be abided by the service provider such as ensuring no mistakes occurs during the ordering process between a customer and the waiter/waitress [20].

2.5 Restaurant Management Information System (RMIS)

Reduce Customer Waiting Time at Point of Sale Restaurant Management Information System (RMIS) is widely used in fast food and beverages industry in the past decade. This phenomenon has inspired restaurant entrepreneurs to also invest in technology due to its great success worldwide. Restaurant entrepreneurs are competing with one another in implementing these emerging technology advancements to get the competitive edge in the market [21]. As stated by researcher [22], RMIS is the core technology used for modern restaurants to effectively treat their customers at the point of sale and assists their employees in managing transactions and inventory control. Moreover,
by utilising RMIS, human errors are able to be mitigated especially during order calculation and order financial transactions. RMIS can impact the growth of a restaurant as it centralised the operation of the restaurant. In addition, this will surge profitability and enable a better environment for customers to dine and eat. RMIS is able to enhance food quality, service quality, inventory control, data transparency, enable smooth operation flow and the capability to quickly acquire financial performance of a restaurant business [22]. Hence, every restaurant must strategies a way to systematically take orders, deliver those order details to the kitchen and charge the customers for their orders. This is possible by implementing an architecture such as RMIS which handles all of those tasks effortlessly and with ease.

2.6 Relationship between Voice Recognition and Natural Language Processing

Speech is a part of human’s everyday life from birth till their dying breath. Speech is developed at a young age for kids through the process of speech communication. Speech comes natural and is the most useful way to communicate with humans. However, the complexity of speech is not realised unless venturing into the perspective of developing a speech recognition system for different languages [23]. Voice recognition is said to be able to identify the speech that is being spoken by a user and to record the speech samples to be analysed. Voice identification converts the speech said by a person that is digitalised to be understood by computers. The converted speech will be in the form of audio samples which will be analysed to understand its meaning in which natural language processing plays a role in it [24].

Natural language processing is a study of interactions between human languages and computers by performing complex computational algorithms to produce meaningful analyses. Hence, natural language processing is utilised by voice recognition to perform a deep understanding of human speech by conducting various techniques to acquire the grammatical structure and the meaning of the speech. Therefore, a useful task can be performed based on the analysis found in the context of the problem to be solved. Natural Language Processing can be used to identify plagiarism, assist computer-aided systems, and the ability to provide better interactivity through speech as compared to using a system manually [25].

2.7 Evaluation of Different Edit Distance Algorithm

The Edit Distance (also known as String Distance) is a prime example of the concept of string matching that is used to measure the distance of strings which originates from the field of computer science. The edit distance is referred to as the evaluation of two string values to identify the distance of two string in regards of how different or similar are the two string values. Moreover, with the assumption that X and Y are two string values, the edit distance is used to measure the minimum number of characters that are required to change the string X into Y by performing insertions, deletions, and substitutions (the term transpositions is also applicable). The edit distance concept was proposed originally for typographical and spelling check but later was used to solve more complex problems [26]. There are numerous algorithms that utilise the concept of edit distance to solve everyday problems. The author will further discuss the string distance algorithms that are relevant to the author’s research which are Hamming Distance, Levenshtein Distance, and Jaro-Winkler Distance.
Hamming Distance is the simplest form of edit distance algorithm. The hamming distance excludes the computation of insertions and deletions when evaluating two strings. As hamming distance only allows substitutions, there is a limitation in ensuring the two strings that are evaluated must be of the same length. Hence, hamming distance measure the minimum number of substitutions which is required to convert a string to be identical to another string [27]. In simpler words, the hamming distance counts the number of characters that are not equal and return an integer value indicating the number of differences as shown in Figure 1. Moreover, the hamming distance algorithm is case sensitive and will substitute characters of the same value but dissimilar in the characteristic of either being uppercase or lowercase.

\[
D_H = \sum_{i=1}^{k} |x_i - y_i|
\]

\[
x = y \Rightarrow D = 0
\]

\[
x \neq y \Rightarrow D = 1
\]

Figure 1: Hamming Distance Algorithm [27]

Levenshtein Distance is a string metric that measures the changes required for two strings to be equal by identifying the number of single-character edits which compromises of three types of edit operations which are insertions, deletions, and substitutions [29]. Each operation is independent of one another with their calculation. The cost for the operations of insertion and deletion is set to 1. The substitution of two characters are also set to 1; otherwise, set to 0. As shown in Figure 2, the minimum bound shows three elements as follows; (1) the first element used for deletion, (2) the second element used for insertion, and (3) the third element used for substitution. The maximum bound uses Hamming distance if the two string sizes are of the same length; otherwise, Levenshtein distance is calculated [30].

\[
D(i, 0) = i, \quad \text{... (4)}
\]

\[
D(0, j) = j, \quad \text{... (2)}
\]

\[
D(i, j) = \min \left\{ \begin{array}{c}
D(i-1, j) + 1 \\
D(i, j-1) + 1 \\
D(i-1, j-1) + \begin{cases} 1 & \text{if } x(i) \neq y(j) \\ 0 & \text{if } x(i) = y(j) \end{cases}
\end{array} \right\} \quad \text{... (3)}
\]

Figure 2: Levenshtein Distance Algorithm [28]

Jaro-Winkler Distance is an improved algorithm that was based on the initially developed Jaro algorithm. The Jaro algorithm is used primarily for name matching. Jaro takes insertion, deletion, and substitution into consideration when evaluating two strings. As shown in Figure 3, the Jaro algorithm calculates the number of \(m\)
which represents the number of characters that are similar; meanwhile, $\tau$ represents the total number of substitutions made which is then divided by two. By emulating the algorithm as shown in Figure 3 and passing the value of $m$ and $\tau$, the Jaro similarity can be achieved [32].

$$d_w = \begin{cases} d_j & \text{if } d_j < b_i \\ d_j + (\ell p(1-d_j)) & \text{otherwise} \end{cases}$$

Figure 4: Jaro-Winkler Similarity Algorithm [31]

Furthermore, Winkler uses the Jaro algorithm as a subset of its algorithm as shown in Figure 4. Jaro-Winkler is a string metric used to measure the edit distance between two strings. The difference between Jaro-Winkler and Jaro is that Jaro-Winkler place more emphasis on the difference at the start of the string as oppose to the end of the string. The reasoning behind the improvement is based on an empirical study that determined that the beginning of a name is where mistakes are more likely to occur. Therefore, as shown in Figure 4, the $p$ is used as a weighting factor (Winkler has provided a default value of 0.1) and $l$ is the length of common prefix which cannot exceed the value of 4. In addition, the multiplication of $l$ and $p$ cannot exceed a value greater than 1. As a result, to obtain the distance of the Jaro-Winker is the equation of $1 – \text{Jaro-Winkler similarity}$. Hence, the lower the value of the Jaro-Winkler similarity will lead to higher the value of the Jaro-Winkler distance, inferencing the two strings are more similar to one another [32] [31].

Overall, hamming distance is used to calculate the number of substitution needed for two strings to equal with the restriction that they are both of equal length, Levenshtein distance is used to calculate the number of character edits to convert one string to be equivalent to a different string, and Jaro-Winkler distance is more suited for short strings especially names to measure the edit distance as it takes account the importance of the beginning of the string compared to its ending. According to a study conducted by researcher [33] that compared various personal name matching techniques which includes Levenshtein distance and Jaro-Winkler distance have concluded that Jaro-Winkler distance is twice the speed of computation compared to Levenshtein distance.

Henceforth, the author has decided to use Jaro-Winkler distance to be used in the system natural language processing feature to determine the dish name being requested by the customer. The dish name that is requested will be evaluated by weighing against the corresponding data available in the database in which the string with the highest Jaro-Winkler similarity will be returned. Additionally, Jaro-Winkler distance is also relatively faster than Levenshtein distance for better computation time which is crucial when handling multiple customers’ requests through the voice recognition feature. On the other hand, Hamming Distance required two strings of the same length which is not appropriate for the context of the problem to be solved.

### 2.8 Evaluation of Different Phonetics Algorithm

Phonetics are algorithms which are developed to identify similar sounding words when pronounced as well to index words based on their pronunciation. Phonetics is used to evaluate speech of human languages rather than the alphabets or letters. Phonetics algorithms are widely developed to be focused on the English language [34].
Phonetics algorithms are designed in detail by taking consideration of all the rules and omissions that are abided by a language. The purpose of evaluating different phonetics algorithms is because each algorithm has their own set of procedures followed to identify similar sounding words. There are three categories of phonetics which are articulatory phonetics that focuses on the sound produces by human organs, acoustic-phonetic focuses on human sound properties, and auditory phonetics that focuses on how the listeners perceive the audio in their brain [35]. The developer will further discuss the phonetics algorithms that are relevant to the developer’s research which are Soundex and DoubleMetaphone algorithm.

Figure 5: Soundex Translation Table

Soundex is one of the earliest algorithms that has stood the test of time. Soundex algorithm will produce a 4-characters string for a word to be evaluated. The first character of the string will remain the same as the first letter of the word. The following three characters will be encoded and reflected upon the Soundex translation table as shown in Figure 5. The other three characters will be appraised based on its phonetics characters and will produce a value such as ‘p360’ for the word ‘peter’ [36]. One of the disadvantages of Soundex is the initial letter of two words or strings to be evaluated that are different will never match [37]. Therefore, words such as ‘Kobe’ and ‘Cobe’ will not match even though they have the similar sound when pronouncing them.

Double Metaphone is an improvement from the original Metaphone that was developed in 1990 that consists of various errors and bugs. The reasoning for why it is named as Double is because this algorithm returns a primary and a secondary code which is useful for unambiguous strings to acquire more accurate values. DoubleMetaphone evaluates a group of letters (called as diphthongs) as oppose to Soundex that evaluates letter by letter [36]. Overall, the DoubleMetaphone has a better design in general as compared to Soundex due to DoubleMetaphone simpler mapping approach. Moreover, DoubleMetaphone was measured to be a bit faster as compared to Soundex based on a study conducted by researcher [33]. In conclusion, the developer will use DoubleMetaphone as the main algorithm for deducting similar sounding words for numbers. For example, strings such as ‘tree’ and ‘three’ must return a match as both the string values sound similar when pronouncing them. However, Soundex will still be used as a backup to capture any similar sounding words that were overlooked by DoubleMetaphone if any.

III. INTELLIGENT SPEECH RECOGNITION SELF-ORDERING MANAGEMENT SYSTEM FOR RESTAURANTS (INTELLIORDER)

The proposed IntelliOrder system will allow efficient management of the ordering process.

- To implement the most optimum ordering process workflow to be used for the development of the software system.
- To implement voice recognition feature for restaurant customers to make an order by speaking their native language that will be translated into the system default language.
- To translate the restaurant’s menu into the desired language of the customer choice.
- To implement a splitting order feature for restaurant customers at the same table to make payment separately to the cashier.

**IV. SYSTEM ARCHITECTURE**

The system functionalities are separated into six target users of the system which are the customers, cooks, cashier, server, manager/owners and administrators. Therefore, each target user derives their own module in the system. The diagrams below (Figure 6 - 11) depict the Data Flow Diagram (DFD) Level-0 of IntelliOrder.

![Figure 6: IntelliOrder DFD Level-0 Part 1](image)

![Figure 7: IntelliOrder DFD Level-0 Part 2](image)
4.1 Customers Module

The customers are able to register for an account that is required of the customer to be able to log into the system to access all the other customer functionalities. The customers can view their promo code in the landing page where they are introduced about the restaurant, and some relevant images are shown. The customers can navigate from one page to another by clicking on the different buttons at using a bottom navigation bar. The customers are also able to
browse the menu on the same page where the customers are able to place an order. The customers can freely look through the menu and add items of their liking into the order data table to be sent later. The customers are also able to view if a menu item is vegan-friendly, look up its ingredients, and macronutrients. The customers are given the option order by speech which is later computed and added into the order data table if valid. After the customer has submitted their order, the customers will be asked to split the order. If they proceed with splitting, they will be navigated to a different page where they can choose the number of seats to split and which menu item goes where before submitting it once last time. The customers can view their order and each of the order items status from start to finish in real-time. The customers are also given the leverage to cancel their order if no order items have been initiated yet by the staffs. After the customers are done with the customer’s stay at the restaurant and all order items have been served, the customers can check-out and proceed to the counter to pay off the bills either as a whole or separately. The customers are also able to view their past transaction for past orders.

4.2 Cooks Module

The cooks can view all the orders made by the customer that is filtered based on their role. The cooks will be updated with order details in real-time. The cooks can update an order item status to indicate that particular cook will prepare the order item, henceforth, all the other cooks will not be able to view the order item that has been assigned to a cook beside themselves. The cook that will prepare the menu item can once again update the order item status to indicate the order item is ready to be served.

![Figure 13: Cooks’ Dashboard Page](image)

4.3 Server Module

The servers can view all the order items prepared by the cooks which are updated in real-time. The server can update an order item status to indicate that particular server will serve the order item to the customer’s table; henceforth, all the other servers will not be able to view the order item that has been assigned to a server besides themselves. The server that will serve the menu item can once again update the order item status to indicate the order item has been served and received by the customers.
4.4 Cashier Module

The cashier is in charge of looking up the table number where the customers had their order taken to be paid. The cashier can select the seat number to pay for an order separately or as a whole (depending if the customers choose to split an order or not). The cashier can proceed with the payment by cash or card depending on the customer choice. After the customer have successfully paid for an order, an e-receipt is sent to the customer’s account email. The e-receipt contains all the order details and relevant information.

4.5 Manager/Owners Module

The manager/owners can view the total sales for the current day, the past seven days, the current month, the current year, and every year of business operation. The manager/owners can view and search staff activity logs that
indicate the check-in and check-out time for every staff that uses the system. The manager/owners are also able to view and search the customer feedbacks. In addition, the manager/owners can view the average score for the ratings.

![Manager/Owners’ Customer Feedbacks Page](image1)

**Figure 16: Manager/Owners’ Customer Feedbacks Page**

### 4.6 Administrator Module

The administrator can view, search, add, modify, and remove details for the main categories, sub categories, staff accounts, and promotional code. Additionally, the administrator is only able to view, search, and modify the macronutrients details.

![Administrator’s Manage Account Page](image2)

**Figure 17: Administrator’s Manage Account Page**
V. CONCLUSION

The final product developed achieved all the targeted goals and objectives of the author for this project. This paper has covered the design and architecture of the IntelliOrder system inclusive of its functionalities. IntelliOrder simplifies the way orders are taken and managed along the ordering process. The system relies on mobile devices such as tablets to operate and function. Therefore, the cost of hiring employees and resources needed can be saved due to the automation process of the system which requires one time purchases for devices that can be used for a long-term duration. Furthermore, papers as the primary medium of order tracking mechanism for the operation of a restaurant can be eliminated and replaced with IntelliOrder. Thus, employees are more content with working in a greener environment which leads to greater productivity and efficiency.

REFERENCES


