

[How to Convert User Lip Movements to Voice Command]

By: Imad Mohammed Elsayid Abdalla.

Assistant professor.

Chief of General Programs - CDC, College of Information Technology and Computer Science,

University of Holy Quran and Islamic Science - Sudan.

Abstract

The science of reading lip movements is processing significantly, such as getting to know character, sentences, phrases from lip movements, the problem of the research is how to convert User Lip Movements (ULM) to Voice Command (VC), and the camera used in interaction must accelerate the speeds of taking pictures to control of continuous taking of ULM and tongue, also the confusion system in discrimination any ULM or overlap one picture into another picture, the research objective convert the ULM to VC, The methodology of research is describe the ULM system and VC in ULM Treatment System (ULMTT), and collected data for analysis phase as dataset, in practical used python to build the model through Viola-Jones Algorithm, and deep neural networks, we follow "DETPCA" converter model in research, the result of research through the processing ULM we have alternatives picture for each different ULM, a certain ULM can be displayed for the user to choose the correct ones as a helping process for user in interactive system, we build converter Model of ULM to product Speech.

key words: Usability Testing, Human and Computer Interaction, Human Computer Interaction Design, Human Behavior Analysis, Experience Design, User Lip Movements, Voice Command.

المستخلص

إن علم قراءة حركات شفاه المستخدم يشهد تقدم ملحوظ، من التعرف على الحرف، الكلمات، الجمل والعبارات من حركات شفاه المستخدم، مشكلة البحث كيف يتم تحويل حركات الشفاه المستخدم إلى أوامر صوتية، والكميرا المستخدم في التفاعل يجب أن تعجل سرعتها في أخذ صور للتحكم في إستمرارية أخذ حركات شفاه المستخدم، وأيضاً إرتباك النظام في تمييز أي من حركات شفاه أو أن تتداخل واحدة من الصور مع صوره أخرى، أهداف البحث تحويل حركات الشفاه إلى أوامر صوتية، منهجية البحث وصف نظام حركات الشفاه و الأوامر الصوتية في نظام حركات الشفاه، وجمع البيانات لمرحلة التحليل كقاعده البيانات وهو نموذج تحويلي في البحث، وفي DETPCA، كما إتبعنا Voila-Jones معيارية، وإستخدمت البايثون لبناء نموذج عبر خوارزمية نتائج البحث من خلال معالجة حركات الشفاه لدينا بدائل صوريه لكل حركات الشفاه المختلفه، مجموعته من حركات الشفاه يمكن عرضها للمستخدم للإختيار الصحيح منها كعملية مساعده للمستخدم في نظام التفاعل، بنينا نموذج تحويلي لحركات الشفاه إلى كلام.

الكلمات المفتاحية: إختبار قابلية الإستخدام، تفاعل الإنسان و الحاسوب، تصميم تفاعل الإنسان و الحاسوب، تحليل سلوك الإنسان، تصميم الخبرة، حركات شفاه المستخدم، الأوامر الصوتية.

1. Introduction

The Usability is interaction objectives, its importance, because the user needs to understand how to be effective in processing with the computer.

The change now is to shift of interaction to speech recognition in open environments, which is based on reading the user lip movements, and to accomplish certain tasks with high efficiency and reliability.

In this study we will focus on extracting the VC directly from the ULM, as the user is speechless, and we want to interact through the VC in the current interactive process. Also, the method used in the interaction does not affect the level of general performance in the completion of operations on the proposed interactive system, since this area is fertile with respect to future studies that can be done in reading of ULM, which will raise the level of usability in the near future.

We want to reach in this mode (lip-to-voice) of interaction to the performance that to be executed with the lip reading that are converted to VC inside the computer.

In the near future, VC can be encrypted with a *verbal gesture* and a new feature in which commercial hardware companies are looking to include these technologies in their software products.

It can also be done in the event that the user cannot hear, to interact through ULM produced by the computer to interact with other users or device.

2. Research problem

The problem of the research is how to convert ULM to Voice Command, and questions

Question 1: NL is best way to interaction of electronic devices?

Question 2: VC will get full success in interaction methods?

3. Objective

The main objective is to produce speech/VC from ULM in human computer interaction system.

4. Research Methodology

The converter model of research is “D” Define of ULM, “E” Extract Pictures of ULM, “T” Treatment of ULM, “P” Picture processing, “C” Configure letters, words and sentences from captured Pictures, “A” Automatically Output the produced Voice Command, in brief its “DETPCA” converter model, and an implementation of research used the Viola-Jones Algorithm to treatment of ULM, and deep neural networks of user’s command.

5. Related studies

The study in Dhiren (2015), the needs to new interaction technologies to support user tasks, provide better access to information, create more robust communication formats, the tools used to design, build, test and evaluate user interfaces, and the processes that developers follow when creating interfaces, the UI study provides a two-pronged way to understand how humans and machines interact, and in the results we gain to understand how the human mind processes information, we get

insight into how human memory treats the information provided, as well as the limitations of it, the recommendations advances in the computer industry and human psychology, we only know the basics of human, if we can understand human better, we can make better interaction designs, new areas such as AI and virtual reality open new doors for human computer interaction, devices new communication like wearable clothing etc. is the future of HCI.

Increased use and acceptance of human computer interaction enhanced continuous development, as a result of the large variety of HCI tools available improving our ability on accessing and using user information, the objectives of paper provides a review of these groups and illustrates the basic barriers, in the paper we describe the aspects outstanding similarities, differences, and what is missing, the result suggest how ideas can contribute to improving human computer interaction in education and practice.

In UK (2002), the paper objectives present a new approach to modeling interaction based on the concept of information sources, which is inspired by a distributed cognitive approach, there are of course many approaches to HCI modeling and the motivation behind this paper is not to offer another method instead, the motivation behind us is that it is clear that recent developments are related to HCI modeling and design, so far, ideas have lacked vision in the HCI community by providing a model of concepts rooted in the concepts of realizing this vision, identifying the resources needed to act as a fundamental axis of interaction between people and technologies, but it has stopped, on providing a definition of such resources at a level that can be used for interaction analysis, the resource model described in this paper identifies a limited number of resource types such as abstract information structures that can be used for interaction analysis, and the result illustrates how these abstract types can be represented differently in an interface, the resource model served the concept of interaction strategy to describe the way in which configurations can be made.

The result illustrates how these abstract types can be represented differently in an interface. The resource model uses the concept of interaction strategy to describe how different resource configurations can shape user actions in a different way, these two components of the resource model, information structures and interaction strategies, through the process of coordination and integration, provides a link between devices, representations and actions that have not been well expressed before.

Adriana, Anna, Simone and Tayana, in Brazil (2015), the increasingly interactive systems exist in everyday life, but many people still have difficulties in using them, we believe that the use of models to systematically represent interaction during systems design may prevent such difficulties.

The study to enhance understanding of user objectives and designer reflection on alternative solutions and decisions related to interaction, the result has not found evidence of the effects of their use on quality and therefore, this paper presents an empirical study on the common use of schemes the interaction of MoLIC most respondents considered creating mockups based on MoLIC charts useful, in addition, the number of defects found in MoLIC diagrams indicates the need to develop techniques for assessment by proceeding with the design process.

Christopher, Jeff, the research report on gestational control of digital music, currently performed at the University of California, Irvine, and is currently performing experiments with various motion performance capture systems, including single video camera systems and 3D optical motion capture system, the result is a series of presentations from different approaches to motion analysis and the assignment of voice gestures.

Ramanana, Gauthier, Bolduc, Berry, Alain, in Canada (2014), the goal is to investigate the reproduction of industrial voice environments or compositions, the method is based on a combination of microphone array recording, beam formation, wave field tuning (WFS), and equality Bass, this approach relies on fixed-direction direction finders to separate signals according to their incoming direction, then the beam signals are processed by WFS in order to re-create the voice environment as multiple plane waves, a parametric theoretical study is provided based on a number of beam orientation directions, resulting in comparisons between target voice environments and environmental clones from in order to determine the appropriate system parameters to reproduce voice environments as described in industrial or composite contexts.

Leo, Symeon and Ville, in UK and Finland (2017), the paper goal is to implement a voice camera program, which uses a spherical microphone array and a spherical camera software that builds on the Cross-Pattern Coherence (CroPaC) spatial filter, the result proved to be Effective in noisy frequency field voice conditions and is based on the identification of cross-spectrum between two synchronous beams, this technique is exploited in this work to capture and analyze voice scenes by estimating a parameter similar to the probability of voice appearing at specific locations.

In this work, we suggest an additional algorithm to suppress side lobes based on the product of multiple packages of CroPaC, the plugin for Virtual Studio Technology (VST) was developed, for both time zone microphone signal conversion in spherical harmonic field and voice camera.

Satoshi, Yuuki, Hiroshi, Koichi, Hirochika, in Japan (2014), describes the paper matrix microphone that can be used to localize voice and capture voice, the system has a sensitivity to all directions, and can act as a remote microphone for the right direction voice is captured by the microphone group by the previous delay and unanimity package, the voice pressure distribution simulation is displayed in a circular microphone matrix of 23 and 64 inches, according to the result simulation, a 32-channel fire wire is developed with a maximum sampling rate of 11 kHz sample , the matrix system is evaluated microfiber Ring 32 inches using voice frequency component.

Omolaro in 2015, the goal of computer systems security mechanisms is to prevent or detect threats or avoid from any eventual attack, the problem in this approach traditionally, this study is an attempt to change the traditional approach by modifying two phases of the computer system design model to integrate a range of use as well as security requirements, this study led to the establishment of a scalable computer design methodology for use and safe, it has been tested using an online questionnaire, data from the previous study were sorted using a synthesis matrix, and analyzed using qualitative content analysis, including some prototypes and prominent methodologies in design, the

results are designed to design usable and secure computer systems to prevent malicious groups or people from stealing organizational information assets or disrupt organizational processes, it was discovered that the reason for the failure of security mechanisms is that it is often used incorrect or not used at all, specifically, most users find that the security mechanisms on their computers are too complex and prefer not to use them, as recommended by respondents who filled out the online questionnaire, implementation of the methodology may require additional financial and human resources.

Christoffer in 2015 showed, the objectives of this paper to provides the development of a conceptual framework for user experience user response design (URUXD) to enhance the link between user by enabling a more relevant and personalized user experience, the tools and methodology that transforms existing ways of adapting to the interface responsibility for designing the user experience resulting from the designer to the user or system, furthermore, we argue that as long as computers cannot translate the meaning of what humans communicate, fast web design responds differently from the concept of user response in the way that the user does not actually respond but to the technical device used, and so the framework enables the integration of multiple designs into a graphical user interface, (GUI) in an information system that responds to the user without risking a violation of the principles of use.

Mladen, in Sweden (2016), targets with many operators monitoring and taking action when necessary, traffic control system operators use keyboard and mouse to interact, to increase operator efficiency and productivity in to control traffic, these gestures are intended to solve specific usability issues discovered during the interview with the user, and aim to answer the following question: Can gesture-based interaction solve usability issues and create simple use the current CTC system, consists of usability issues and requires improved user interface interaction, operator setup, tools and methodology in interviewing system users and researching related topics leading to a new design coupled with interaction and user interface as well as some suggestions to increase productivity, conduct exploratory research on a topic involving the design, implementation and testing of hand movements with users, and display gesture-based interaction, the results are simulated for traffic control systems that tend to improve the process, many gestures such as panning, zooming, zooming using the map are designed using the Leap Motion controller allowing for intuitive interaction, from a comfortable perspective, the body position and hand position of the operator are examined and the suggestion of using workstations in order to reduce pain and discomfort at work, laboratory tests confirm that gesture-based interaction achieves more natural and intuitive use of traffic control systems, the recommendations there is a high probability of gesture-based interaction to increase usability achieving effective control of operators would reduce train delays and maintain a safe flow of traffic.

George in 2017, aims our work to create new capture and analysis techniques and new scenarios for input using keyboard and mouse, tools and methodology, using the keyboard and mouse, the results believe that HCI behavior analysis can be used to infer cognitive states, the recommendations presented on these pages represents only the first steps in understanding users better by analyzing their HCI behavior.

Christian in in Hagen's state (2016), aims to develop the theory of emotional communication in order to explain and understand the process of taking a perspective in human interaction on the pre-reflection body and thus the emotional level, tools and methodology develop the theory of emotional communication in order to explain and understand the process of taking perspective in human interaction on a physical level, pre-reflection and thus emotional, the dominant side acts as a perspective donor and the sub aspect acts as the future, the dominant side conveys its perspective to the sub side, any transition of perspectives within the interaction based on physical tendencies or forces gravitational and repulsive, this is consistent with the results in studies that have a positive impact on perspective taking, the dominant side conveys its view on the sub side, and in this sense, human interaction is always about conversion, negotiation or conflict of perspectives based on the works of Schmitz in 1964.

The study was conducted by Seoul National University, this study aims to develop a task-based usability checklist based on indicative assessments of mobile user views, tools and methodology develop a hierarchy of UI design elements and principles of use related to mobile phones and then use them to develop a checklist to explain the practical effectiveness of the proposed checklist, there is a growing demand to explore appropriate assessment methodologies that measure the usability of mobile phones quickly and comprehensively, most of the usability problems found through usability testing and additional problems were discovered through the proposed checklist, on the most commonly used evaluation methods.

Hani, Muhammad, Farah, in 2014, the main objectives of this study is to measure the impact of the quality of mobile applications perceived by the user, and mobile features on the continuity of mobile shopping by perceived benefit, the methodology of quantitative data collection is applied to gather appropriate data, and uses multiple hierarchical regression analysis (HMRA) to test both the study's hypothesis and model, and the results reveal that the quality of mobile applications the user sees affects to act on the intention of continuing mobile shopping through perceived benefit, and trust.

The recommendations that operators should improve the radio coverage and speed of 3G data transmission in all regions, as this seriously affects the experience and experience of customers, mobile production companies should focus on improving mobile phone features that have shopping process such as weight, size and screen display, these factors affect the ease of shopping, and from another point of view.

Najmeh, Mahboubbeh, Abol, Vahid in Iran (2015), the objectives of paper to compared some of the methods of conducting the usability test most commonly used in human computer interaction, interfaces, designers can enhance usability by testing and optimizing the current interface, designers can also evaluate the possibility of using the current interface, borrow their own strengths, improve their shortcomings, apply them in the new design.

6. Analysis of User Lip Movements Treatment System (ULMTT)

The ULM processing system consists of the definition of lip movements, treatment and from the camera to take instant pictures of the lip movements as well as the processing of pictures taken,

including the formation of letters to words and words are installed, sentences and finally generate the voice section, see figure (6.1).

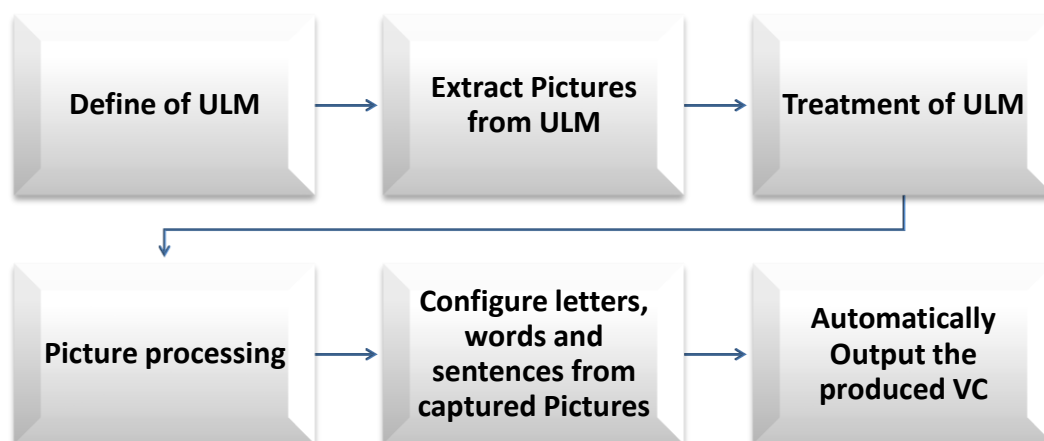


Figure (6.1): a Converter Model of UML to VC

6.1 Define of User Lip Movements

The ULM of the user has connotations understood by humans even in the normal communication between two people in a subject, but there are some gestures of the ULM have their own meaning, so we build dataset for user lip movements meaning, also understood by the recipient of the voice gesture and result in some important information in the process of dialogue between two people, but if these signals are lost in the interaction, it loses some meaning, what we can do?, so 3 dimension model to give more control on it, and the information that is supposed to pass (to another communication side) to the future of the gesture or the lip

movement, but if these signals are lost in the interaction, it loses some meaning, what we can do?, that is the information that is supposed to pass (to another communication side) to the future of the gesture or the ULM, see figure (6.2).

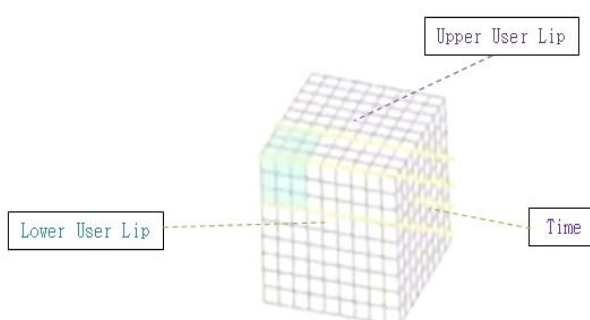


Figure (6.2): a 3 dimension model to Define upper, lower, and time of ULM.

The concept of the connotations of the ULM is hot now, in the sense that they are often used to convey information very important to the other party involved in the dialogue, for example used in social networking, as well as the frowning face with the inclusion of lip inside a sign or anger, the face holds the upper lip far away from the lower lip indication and expression of the state of surprise, all of these precedents are great importance in the process of human computer interaction(HCI) in the

interactive system of the ULM used in the completion of tasks underdeveloped effectively in the interactive process.

I would like to point out here that the ULM can be extracted from the letters and words and even the structures of sentences that serve the interactive process directly, an example the inclusion of upper and the

lower lip for a short period indicating that the letter to be spoken is the letter "O", as well as the movements of the lower and upper lip, so that they have a simple break with the flat tongue movements to the upper gums, look at figure(6.3).

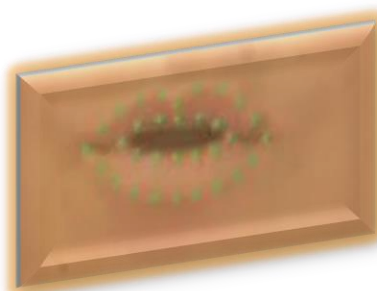


Figure (6.3): Define of upper, lower, and time ULM.

6.2 Extract Pictures from User Lip Movements

The camera found to capture the instant pictures of the ULM must be characterized by high speed, because ULM changing quickly, smoothly

and accurately so you must. distinguish each ULM, which in turn means a particular character from the set of characters sequential followed by others, see figure (6.4).

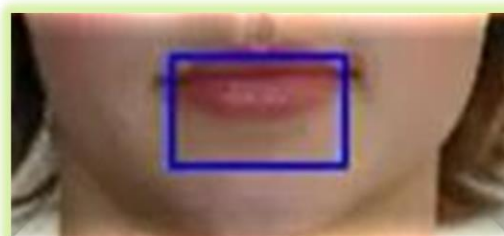


Figure (6.4): Pictures of ULM.

In the real-time the pictures captured, the camera should be a set of sequential pictures in a particular order that corresponds to the sequential ULM at certain interactive moments.

The reaction camera also includes a series of pictures sequenced in the order of the previous lip movements, which is characterized by efficiency, accuracy and speed in the completion of the interactive process.

All barriers that limit the effectiveness of the lip capture through camera should be removed, such as a barrier between the camera and the ULM camera used, for example, the mask, a barrier that reduces the level of clarity of the captured pictures, such as dust, or low visibility in the reaction environment, interaction the night for example.

It also requires the camera to capture the ULM to distinguish the distortions that appear on the lip of some users in the interactive system that does not affect much of the interactive process, but that the system includes special treatments for a part that carries congenital deformity by providing suggestions for the form of the supposed movement, as well as users who suffering from the problem of stuttering causes a slight interruption in the sequence of lip movements that system he take into account.

The lip movements system should have a second night camera for problem solving, low light in some reaction situations, or motion sensors focused on lip motion and gesture.

It is a sign that the letter to be spoken is the letter "w", as well as all the ULM are traced to be the letters of these characters were grouped together to form words, and then the synthesis of this set of words to be the sentences and phrases.

6.3 Treatment of User Lip Movements

The process of handling the ULM requires a wide base of pre-stored ULM to be compared in case of confusion of interactive system of motion

recognition, and therefore followed by confusion also in the meaning and significance of the captured motion.

It is also possible that there is an assistant operation from the user of the system, meaning that the proposed movements and characters that are mixed with the system are distinguished, for example the lip movements involved to any character follows from this set of characters, and the assistant here from the user is to remove the confusion by choosing the correct character of the ULM, see figure (6.5).

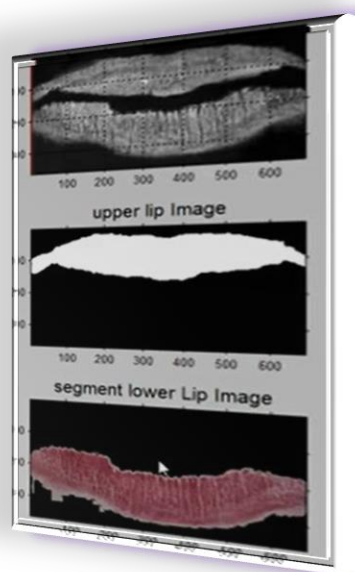


Figure (6.5): Treatment of ULM.

I would like to point out that the user does not interfere very much in distinguishing the ULM with the letter corresponding to the movement, because this negatively affects the effectiveness of the treatment system of the sanctity of the lip, but the intervention here is true to critical and extreme cases.

The lip movements system learned during the interactive process to find the right letter for the particular movement, as well as the formation of the correct words, which directly emanate phrases, sentences and even interactive paragraphs, see figure (6.6).

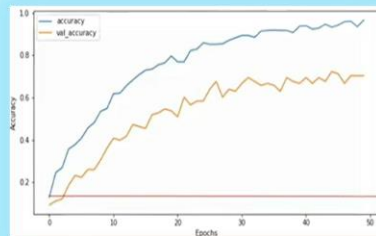


Figure (6.6): accuracy of ULM captured.

In the interactive system, the soft system must create a huge set of words on which to compare sequential ULM at a certain frequency, phrases are constructed on the basis of suggestions from the system to be resolved during the specific interactive process.

6.4 Pictures processing

The process of pictures processing requires high capabilities to extract information of which the significance of interactive value in the implementation of operations to accomplish a specific task for the user and the pictures to be processed here do not differ greatly in terms of the general perspective as pictures of the ULM interactive system, see figures (6.7),(6.8) bellow.



Figure (6.7): character “V” processing of ULM captured.



Figure (6.8): character “P” processing of ULM captured.

The interactive process here requires very fast outputs of the pictures captured at the moment and does not lose the cognitive value only after certain moments have been taken, or be used in its secondary value later in the self-learning of the interaction system, and the precise processing of the pictures taken of the lip movements requires that there are certain dimensions in terms of the horizontal distance as well as the vertical distance, and the horizontal distance represents the length of the lip starting from the far right to the left and the dimensions that mediate it.

The vertical distance represents the highest movements or limit of the upper lip and the lower movements down and down the middle, as well as the movements of the lower lip from the highest movements or limit and down to the lower movements down and in between it and thus, the synthesis of words from the sequence of movements of the tongue and lip together and come up with valuable information serve the interactive

process effectively in the completion of various interactive tasks, including the start be the sequence of sentences that make up paragraphs, see figures (6.9), (6.10).



Figure (6.9): constant picture of ULM

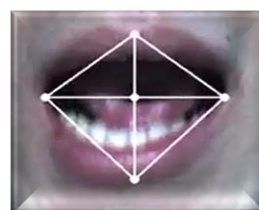


Figure (6.10): picture capture of ULM.

6.5 Configure letters, words and sentences from captured Pictures

for the characters to be formed, there must first be controls, rules and steps to be taken into account during the interactive process, the controls that understand the clarity of the ULM are aimed at such a character, for example joining the upper lip and lower lip of the middle with a little breakthrough in the middle and also the movements of the tongue to the middle with a slight contraction inside with almost flat diastole in the middle of the cavity of the mouth, from this we understand that the pictures taken crosses for one letter that is not similar in this movements is the letter F "and", and so the vetch characters in the Arabic Language and what corresponds in English the same concept.

The rules are that some characters may get confused and here the camera must capture the movements of the tongue within the hollow and the meaning here, that at a moment may be absent from the picture taken movements of the tongue for some reason and here must put a certain rule to separate in the ambiguous characters that this movements follows for such character.

The composition of words is understood only if there is a sequence of a certain lip movements, in other words, if this sequence is lost, the overall structure of the target word is lost, so the synthesis of the pictures sequence of the ULM at certain moments corresponds to a particular word and either in the composition of sentences lies in the separation of words is the key to the formation of sentences or phrases, or in other words, the intervals between the sequential movements of the captured lip is the end of the first word and the beginning of the next word in the phrase or sentence and so on, either commas are not counted in the case of character recognition, any movements the first letter and the movements of the second letter in the word does not depend on the commas, because it causes the loss of the basic structure of the word, while the sentences distinguish between them strokes any long commas so that users can train them.

6.6 Automatically Output the produced Voice Command

This stage is the final stage in the concept of extracting VC or speech recognition of the production directly from reading the ULM during the interactive process, after the ULM was captured by the interaction camera and then interpreted the pictures taken in certain letters, words and sentences, we keep knowing the meaning of the word and its meaning.

The meanings and connotations of the ULM depends on the formation of letters to complete the meanings of the words to the fullest, and must be very careful that the same words that were formed from the ULM captured by the interaction camera and was interpreted the same pictures that carries

different meanings if the word is formed in different movements Arabic, for example, or words that follow the English language may have the same voice but differ in their meaning and letters as well.

Here we want to exclude the previous paragraph some of the following examples, in the sense that there are words that have the same syllable but differ in the meaning of the word such as all, and other words have the same number of letters in the Arabic language, but if the words formed in different movements gave a meaning completely different from the first meaning, for example, the word "book" by breaking the letter K and open the letter T and soothing the letter B this first meaning of the word and the collection of books on the weight of the verb, the other meaning of the same word in the formation of characters in different movements, such as "book" K and tighten the letter T and soothing the letter B to the word and collected clerks on the weight of the act, and another example, for example, the word "teacher" by combining the letter M and open the letter of the eye and break the letter B and the second letter M word, this is the first meaning of the word and collected by teachers, while the other meaning of the same word in a way that the formation of characters in different movements such as "teacher" by opening the letter M and soothing the letter of the eye and open the letter L and the second letter M and soothing the word and collected landmarks and so I priest the rest of the words of the language.

7. The Results

The result of the research are:

- ULM converter model to product Speech.
- ULM systems learned to sort between the user's familiar VC, a program that reads the ULM in order to obtain the correct user's voice commands, this is done by matching the movements of the ULM stored with the captured user VC.
- Converter Mode can read and predict character of ULM.
- The Questionnaires result see the figure (7.1).

Nam e	Sampl es	Confiden ce	Sig	level of Significan ce	Mal es	group (20- 30) Age	Bachelo r's Degree	more than 5 years of Experienc e
valu e	77 sampl e	95%	15 %	.000%	70.8 %	37 F	67.1%	55.4%

Figure (7.1): results of statistical analysis of hypothesis's.

- The hypothesis 1: probability value is 0.000, which means that the majority of respondents agree that interaction by NL is the best way to interact.
- The hypothesis 2: probability value is 0.000, which is less than the 5% level of significance, interacted by VC will get success in the global competition market compared tradition interaction method.

8. Recommendations of Research

They are many recommendations to another researcher in this filed:

- The researcher recommends to more modern interaction methods in order to reach satisfactory of users and be a real benefit to the technological devices of smart methods used in the field.
- There must be various devices technology with new features distinguish them from previous generations, also from the security aspects here is to pay attention to the encryption of conversations lip movements between a computer and another user.
- I also recommend that researchers to the field of intelligent system to sign reading lip movements and linguistic gestures to enable us to match the user's commands.
- Note that this converter model is designed to serve mainly users who are unable to *speak* and other goals regarding the health problem of users and birth defects in humans, take in your consideration human rights.
- Systems can also communicate with each other in a visual way such as readings of movements that appear on the output screens.
- Some issue suspends like, what happen if users aware mask, user talked fast, if user cover the mouth, user have accents.
- More users for training on converter model, and more classes.
- User lip its open book needs to reads in your consideration.
- The models that build must decide what language of user.

9. Conclusion

The voice processing system receives commands only by the user in the form of lip movements that are captured with little change to the user's VC and only in direct interaction environments, without or with repeated verbal commands, but in some cases the voice processing system may receive movements other than the user's movements, in the target environment of the interaction is that other users are in front of the reaction camera, so the set of captured ULM may cause confusion to the reaction system, any of these movements are the command of the target user to interact, and that is in open interaction environments, and the result of the study was that the problems of the interaction of user lip in closed environments were matched by matching the ULM with the movements of the user previously stored, thus execute the VC from the first time issued by the problem of interaction in open environments was solved in two ways, one of which is that the lip treatment system receives commands (ULM) from the user and other lip movements from the interaction environment, then the lip treatment system displays the voice pictures of the user to make the selection process which of these commands is the user's command and after the selection executes the VC, and the other that the lip treatment system matching the movements of the lip with the beginning of the capture of ULM from the open interaction environment can suggest the rest of the ULM or exclude other non-conformity ULM from the first glance.

10. References

The reference of the research is:

Abdou S., Rashwan M., Al-Barhamtoshy H., Jambi K., and Al-JedaibiW., "Enhancing the Confidence Measure for an Arabic Pronunciation Verification System". Proceedings of the International Symposium on Automatic Detection of Errors in Pronunciation Training June 6 - 8, KTH, Stockholm, Sweden, 2012.

Al-Barhamtoshy H., Jambi K., Al-Jedaibi W., Abdou S., and Rashwan M., "Adaptive English Pronunciation Errors for Arab Learners of English", the 12th International Conference of Language Engineering, ESOLE-12, Ain Shams University, Cairo, Egypt, 2012.

Al-Barhamtoshy H., Abdou S., and Al-Wajih F., "A Toolkit for Teaching Arabic Handwriting", International Journal of Computer Applications (0975 – 8887), Vol. 49, No. 23, pp 17-23, 2012.

Anthony I. Wasserman. 'Extending State Transition Diagrams for the Specification of Human-Computer Interaction'. In: IEEE Trans. Softw. Eng. 11.8 (1985), pp. 699–713 (cit. on p. 10).

Ahmed, A.A.E. and IssaTraore (2007). \A New Biometric Technology Based on Mouse Dynamics". In: IEEE Transactions on Dependable and Secure Computing 4.3, pp. 165.

Cooper, A., Reimann, R. & Cronin, D. (2007). About Face 3: The Essentials of Interaction Design. Indianapolis, Indiana: Wiley.

Chris W. Johnson and Michael D. Harrison. 'Using temporal logic to support the specification and prototyping of interactive control systems'. In: International journal of manmachinestudies (1992) (cit. on p. 15).

Carlson, John R. et al. (2004). \Deception in computer-mediated communication". In: Group Decision and Negotiation 13.1, pp. 5.

John Smith, "Web Page Design: Heuristic Evaluation vs. User Testing", December 7, 2010.

A case studies. Informatics in Education, 7(2), pp. 181-210.

Hadjerrouit, S., 2008. Towards a blended learning model for teaching and learning computer programming: A case study. Informatics in Education, 7(2), pp. 181-210.

Hassanin Al-Barhamtoshy, SherifAbdou and Mohsen Rashwan, Mobile Technology for Illiterate Education, Cairo University, Life Science Journal 2014;11(9), 20114, <http://www.lifesciencesite.com>.

Jokela, T. (2004). When good things happen to bad products: Where are the benefits of usability in the consumer appliance market? Interactions, 11, 28–35.

Jamal, H. and Shanaah, A., 2011. The Role of Learning Management Systems in Educational Environments: An Exploratory Case Study, Master Thesis, Linnaeus University.

Johannes Itten, and Faber Birren, The Elements of Color: A Treatise on the Color System of Johannes Itten Based on His Book the Art of Color, John Wiley and Sons, 2003, ISBN 0471- 289292-9.

Kinchin, I., 2012. Avoiding technology-enhanced non-learning. British Journal of Educational Technology, 43(2), pp. E43-E48.

Koehler, M. and Mishra, P., 2009. What is technological pedagogical content knowledge (TPACK)?. Contemporary Issues in Technology and Teacher Education, 9(1), pp. 60-70.

KarstenLoer and Michael D. Harrison. 'An integrated framework for the analysis of dependable interactive systems (IFADIS): Its tool support and evaluation'. In: Automated SoftwareEngg. 13.4 (2006), pp. 469–496 (cit. on p. 17).

Liao, Q., Pan, Y., Zhou, M. X., & Ma, F. (2010). Chinese online communities: Balancing managementcontrol and individual autonomy. In Proceedings of CHI'10, April 10-15, Atlanta, GA, USA (pp. 2193-2202).

Laurillard, D., 2008. Technology enhanced learning as a tool for pedagogical innovation. Journal of Philosophy of Education, 42(3-4), pp. 521-533.

Laura Lin, Kim J Vicente and D. John Doyle. 'Patient Safety, Potential Adverse Drug Events, and Medical Device Design: A Human Factors Engineering Approach', In: Journalof Biomedical Informatics 34.4 (2001), pp. 274–284 (cit. on p. 26).

Mulder, S. &Yaar, Z. (2007). The User is Always Right: A Practical Guide to Creating and Using Personas for the Web, Berkeley, California: New Riders.

Mulwa, C., Lawless, S., Sharp, M., Arnedillo-Sanchez, I. and Wade, V., 2010. Adaptive educational hypermedia systems in technology enhanced learning: a literature review. In: Proceedings of the 2010.

ACM Conference on Information Technology Education. Midland, MI, USA, 07-09 October 2010. ACM. pp. 73-84.

Niedenthal, Paula M et al. (2005), 'Embodiment in Attitudes, Social Perception, and Emotion", In: Personality and Social Psychology Review 9.3, pp. 184{211.

NajmehGhasemifard, MahboubehShamsi, Abol Reza RasouliKenar&VahidAhmadi, A New View at Usability Test Methods of Interfaces for Human Computer Interaction, Global Journal of Computer Science and Technology: A Hardware & Computation Volume 15 Issue 1 Version 1.0 Year 2015.

Paul Curzon, RimvydasRuk's'enas and Ann Blandford. 'An approach to formal verification of human-computer interaction'. In: Form. Asp. Comput. 19.4 (2007), pp. 513–550 (cit. on p. 13).

Reaney, Patricia (2012). Most of world interconnected through email and social media.

Rusche, Bryan and Erin Elofson (2007). The State of Communications Today. Tech. rep. Microsoft Corporation.

So, H. J. and Bonk, C. J., 2010, Examining the Roles of Blended Learning Approaches in Computer-Supported Collaborative Learning (CSCL) Environments: A Delphi Study, *Educational Technology & Society*, 13(3), pp. 189-200.

Sam Owre, John M. Rushby and Natarajan Shankar, 'PVS: A Prototype Verification System'. In: *Proceedings of the 11th International Conference on Automated Deduction: AutomatedDeduction. CADE-11*, London, UK, UK: Springer-Verlag, 1992, pp. 748–752 (cit. on p. 12).

Scissors, L., Shami, N. S., Ishihara, T., Rohall, S., & Saito, S. (2011). Real-time collaborative editing behavior in USA and Japanese distributed teams. In *Proceedings of CHI'11*, May 7-12, Vancouver, BC, Canada (pp. 1119-1128).

Sood, S., Mbarika, V., Jugoo, S., Dookhy, R., Doarn, C. R., Prakash, N., & Merrell, R. C. (2007). What is telemedicine? A collection of 104 peer-reviewed perspectives and theoretical underpinnings. *Telemedicine and e-Health*, 13(5), pp. 573-590.

Song, W., Yu, H., Liang, C., Wang, Q., & Shi, Y. (2012). Body monitoring system design based on android smartphone. In *Information and Communication Technologies (WICT), 2012 World Congress on. Trivandrum: IEEE*, pp. 1147-1151.

How User Experience is Understood? *Science and Information Conference*. London, UK 2013, pp. 791-795. *IEEE Conference Publications*.

Vinciarelli, A., Murray-Smith, R., & Bourlard, H. (2010, September). *Mobile*.

Zhou, X., Sun, S., & Yang, J. (2014), Sweet home: Understanding diabetes management via a chinese online community. In *Proceedings of CHI'14*, April 26-May 1, Toronto, ON, Canada (pp. 3997-4006).