Siri in healthcare: a "Master Class" in artificial intelligence and the future of Apple computing

Posted on 2018-06-04 by Gregory Bufithis in Artificial Intelligence/Cognition

Siri in healthcare : a "Master Class" in artificial intelligence and the future of Apple computing

I. INTRODUCTION

II. "WAIT TIME" IN HEALTHCARE

III. CONCLUSION

4 June 2018 (Washington, DC) – mHealth (mobile health) is a general term for the use of mobile phones and other wireless technology in medical care. The most common application of mHealth is the use of mobile phones and communication devices to educate consumers about preventive health care services.

But as I learned at the **Mobile World Congress** earlier this year, it includes a broad range of innovation and advances in the growth and development of mobile connectivity to support the entire healthcare eco-system. This includes all aspects of mobile medical healthcare products, services, solutions and applications that are emerging today, from de-centralised access to health services and remote diagnosis solutions, chronic disease management and healthcare monitoring to treatment compliance health coaching as well as everyday lifestyle apps for well-being and fitness.

And that is the trick: in a world where we are inundated with artificial intelligence technologies — speech technologies, text processing, Natural Language Processing, contextual computing, etc. — plus devices like wearables, phones and all manner of IoTs, we yearn for an integration of AI technologies and devise into a useful product.

Zaid Al-Timimi seems to have done just that with his ***"The Patient Is In"*** family of apps – in what appears to be the very first product to integrate AI, Siri, and Apple Watch. And he has done it to solve a problem faced by all doctors in small medical facilities and major medical facilities, too: reducing patient wait times by improving the communication of doctors and charge nurses through the integration of secure messaging with AI, wearables, virtual assistants, and Internet of Things technologies.

And by doing so, as you will see below, he has given us a "Master Class" in the artificial intelligence technologies that dominate today's AI world: speech technologies, text processing, Natural Language Processing, contextual computing, etc.

I. INTRODUCTION

A. A few words about Zaid

Zaid Al-Timimi is a technology executive and a former member of the JCP Java Standard Committee and chair of the Object-Relational subcommittee of the ODMG Object Data

Consortium. For more information about Zaid, visit his LinkedIn profile **_here_**.

My regular readers will recognize Zaid's name. In 2012, at LegalTech New York, I used his very cool "The Mashup App" which was a vision for the future of the web: a web where we control our personal information and curate our digital data — from memories to knowledge — from all from our personal devices. (At the time, mine were the iPad and iPhone). The app is a personal database in which you save your digital data. You can save all types of data including web, PDF, images, video, audio, location, and time data. Since your personal database is on your device, you can save everything that is important to you – in this case, a log of every place I visited during my 3 days at LegalTech that year, plus the 3 days after in NYC at client meetings.
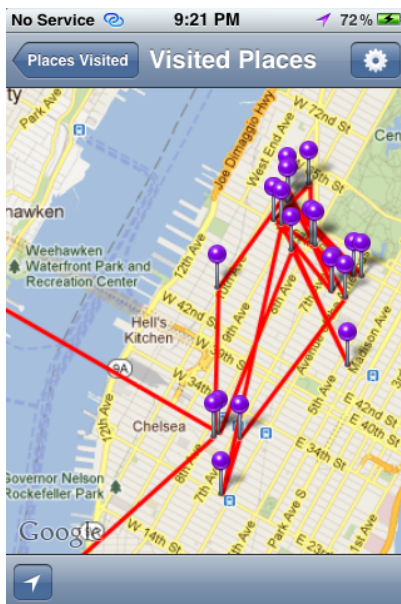
From this I could generate a "diary" of my day, or my week and I could also associate any document (such as a PDF or a restaurant receipt or dry cleaning bill) I had to that "pin" location and/or day.

It is brilliant. After searching or browsing your personal database, you can see your results using 3D visualization, listen to it using text-to-speech, or use augmented reality to project the data onto your device's camera or map. Best of all: you can share your data using email, Twitter, and SMS text messages. And unlike a file name, you can describe the data using full sentences in multiple languages. This helps a person to remember their data. And unlike folders, you can aggregate — or mashup – the data into multiple categories

But the most thrilling part for me was the video mashup. Our video crews shot 25+ videos at LegalTech that year (interviews plus background segments) and we were able to load all of them onto my iPad and create categories containing all the videos as well as accompanying PDFs or other material of the folks we interviewed. I was then able to review, edit and clip from each video the appropriate piece to illustrate my LegalTech analysis via the interface:
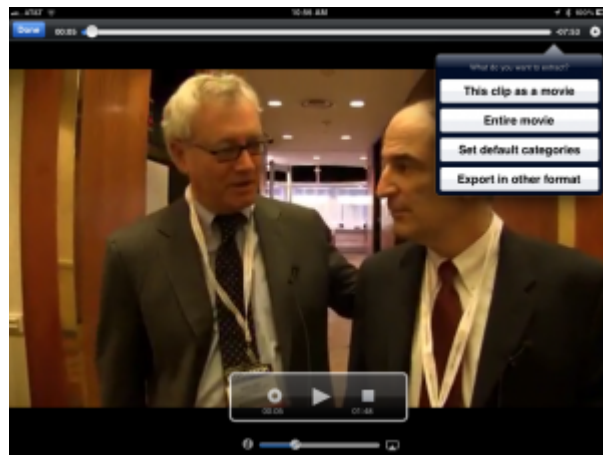
I wrote a detailed post on "The Mashup App" and that incredible 2012 experience which you can read by **_clicking here_**. A number of my journalism colleagues also use it.

This will explain the brilliant mind behind this newest healthcare innovation which I will now describe.

B. _About Zaid's new App_

Founded to improve doctor productivity throughout the world, Zaid launched **_"The Patient Is In"_** family of apps in September 2016. His goal was to solve a problem that scores of doctors told him about, from both small medical facilities and major medical facilities: reducing patient wait times. He did it by improving the communication of doctors and charge nurses through the integration of secure messaging with AI, wearables, virtual assistants, and Internet of Things technologies. Moreover, he did so not in a vacuum but by working with top doctors in Washington DC as well as charge nurses, anesthesiologists and surgery teams in order to full understand the issues — and by doing so extend (and further develop) a desktop solution so that it became a mobile solution.

And even if your interests do not include the healthcare and medical fields, you'll learn why in modern AI labs today speech technologies are considered a core feature in any computer system, and how the three speech technologies which comprise the ever-more-used voice interface work.

Let's get into some of the detail.

II. "WAIT TIME" IN HEALTHCARE

If you ask the question "what do you hate most about going to the doctor's office," many people will say: "it's the wait time". If you ask doctors the question "what do you hate most about treating patients," many will say: "it's the wait time". The patient must wait until the doctor shows up and the doctor must wait until she is told by the charge nurse that the patient is ready to be treated and is provided with any relevant notes about the patient. Those notes may include the patient vitals signs or just may be the room number in which the patient is waiting.

Until then, both the doctor and patient must wait, and wait, and wait.

> NOTE: some interesting bits Zaid learned in his research. Some patients are now suing their doctors for making them wait too long or simply switching doctors. Other patients – such as hourly workers, seniors who depend on caregivers, as well as those who have unreliable transportation such as the handicapped or blind – are more likely to avoid seeing a doctor until their health has declined to a point where they have no choice but to seek treatment. Delayed treatment is always more expensive and stressful. So, reducing patient wait times has both a psychological and economic benefit for patients.

There have been some recent approaches to reducing patient wait times in emergency rooms, but those approaches are related to process improvement such as delegating authority to nurses to order lab work, having assistants rather than doctors fill out paper work, and better patient triage. Zaid went one-step further by developing an app-based solution to significantly reduce patient wait times, building in collaboration with tech-savvy doctors and nurses from all major medical fields.

A. Using AI and Siri

Those of us working in the AI community know the old joke in academia about which technologies constitute artificial intelligence and which constitute plain old computer science: if the professors in the computer science department did not understand a technology then it was considered to be AI, otherwise it was considered to be computer science. Another way of understanding the difference between computer science and AI is that computer science is about finding a correct answer, true or false, 1 or 0, black or white … and AI is more about finding a likely answer, percentages, and shades of grey. During my conversations with Zaid he noted this:

> *Ten years ago, the use of 3D graphics in games was considered to be AI and only a small percentage of programmers knew how to simulate the physics of bullets, objects colliding, and how to create light sources and realistic textures. With the introduction of iPhone and the standardization of 3D technologies, those elite game developers documented their algorithms and licensed their internal software libraries and game engines allowing any app developer to create cool games. That started the App Store revolution and allowed any programmer to become a 3D game developer. Today, popular games generate millions of dollars each day.*

Over the last five years, the AI technologies required to process photos and videos have become mainstream allowing any developer to create sophisticated camera apps, photo filters, and even generate a realistic simulation of a person to create fake movie scenes. Likewise, the AI technologies used to process and generate speech and text content and ultimately understand human language are now becoming mainstream allowing developers to create apps and services which integrate with digital assistants from companies such as Apple.

The Siri digital assistant is available on all of Apple's devices and is a combination of specialized software and hardware. Though primarily accessed by voice, an iOS user may also access Siri with a keyboard. This feature is named "Type to Siri" and is enabled in the accessibility settings. While Siri performs many built-in tasks, iOS and watchOS app developers are able to integrate with Siri and thereby extend their apps with a voice interface. With iOS 11, about a dozen types of apps can be integrated with Siri to provide users with a voice interface. So rather than tapping on an app's icon to launch it and t'

tapping some more within the app to perform a function, Siri empowers you to use your voice to directly have an app perform a function.

And for you iPhone users, to see a list of Siri's built-in features and the additional services provided to Siri by the apps installed on your iOS device, invoke Siri and ask "what can you do". If you have enabled Siri on your iOS device, you should see screens similar to the following:
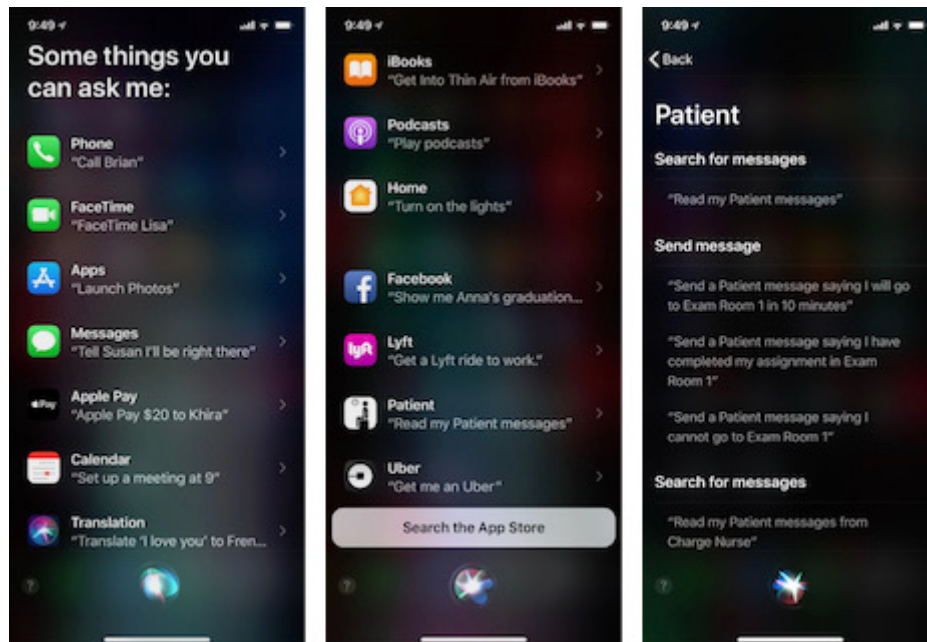
1. Speech Technologies

Created in the AI labs of the modern computing era, speech technologies are now considered as a core feature in any computer system. Originally driven by the need to create accessible systems for the print-disabled and then by the need for foreign language translation, speech technologies have now become a mainstream technology due to the evolution of the modern man-machine interface: from the command-line and graphical user interfaces found in desktop terminals and computers to the touch interface of mobile devices and continuing with the voice interface found in virtual assistants.

There are three speech technologies which comprise the voice interface:

a. speech recognition which converts sound into language,

b. speech synthesis which generates speech from text content, and

c. natural sounding voices representing different genders in different languages with various accents for different locales.

(a) speech recognition

Siri transcribes your speech into words and then passes that text to be processed by an app which you have identified. If you do not specify an app to process your voice request, Siri will attempt to handle the request with its built-in services.

The transcription may be incorrect if you have a thick accent; use uncommon or foreign language words, use abbreviations or technical jargon; or are in a noisy environment. All Apple devices have specially tuned microphones and other hardware and software to help in the speech recognition process. Siri can learn new words by adding them to the iOS dictionary. On its own, Siri can automatically learn that you tend to pair certain words together and will attempt to use these phrases in its audio transcription rather than use a more phonetic transcription.

Simple apps do not need to determine if the audio transcription provided by Siri is correct but mission-critical apps have to determine correctness. For example, the *"Patient Is In"* app will disregard the audio transcription "all go to our two" and instead rewrite the phrase to be "I'll go to OR 2". Here's a screen shot of using Siri with the app for Apple Watch:

It is integrated with Siri on iPhone, Apple Watch, and the new HomePod smart speaker.

(b) speech synthesis

While speech recognition is the transcription of speech to text, speech synthesis is the generation of speech from text using an artificially generated voice. Siri uses speech synthesis to confirm a user's instructions and disambiguate between choices. Zaid provided this example:

> *If an anesthesiologist has just started her shift she can ask Siri on the anesthesiologists' HomePod: "Read my patient messages" and using text-to-speech, Siri will read aloud her assignments: "You have a new Patient message from charge nurse: 'You have an assignment in Recovery Room 3 with the note "patient discharge" and were assigned here 10 minutes ago.'" And when the doctor has finished her assignment she can use Siri to respond: "Send a Patient message saying I'm done" and Siri will confirm: "Your Patient message to charge nurse says: 'I have completed my assignment in Recovery Room 3'". The charge nurse will see on her Patient is in iPad app that the anesthesiologist status is available and can treat another patient and that Recovery Room 3 requires cleaning. This reduces the patient wait time for the next patient.*

Another use of speech synthesis is to announce patient assignments and other messages over AirPlay compatible speakers. Small doctor offices are usually in leased office space and it is too expensive to rip up the ceiling and install speakers. In hospitals, the charge nurse usually cannot make an announcement in a specific area. The Patient is in app offers an inexpensive alternative by providing support for making announcements over WiFi using AirPlay compatible speakers including HomePod, Apple TV, and speakers from other manufactures. As seen in the following screen shots of the Patient is in iPad app, the charge nurse can direct a general announcement or the announcement of a patient assignment to any AirPlay speaker such as the HomePod in the Anesthesiologists' Lounge.

(c) natural voices

Advances in machine learning, smarter algorithms, more storage, and enhanced computing capabilities have allowed Apple's mobile devices to be able to speak in more natural sounding voices. ***"The Patient Is In"*** app harnesses this capability to allow the charge nurse to broadcast an understandable and natural sounding wireless announcement in any language to one more AirPlay compatible speakers. And it supports most languages including Arabic, Chinese, English, French, German, Hebrew, Italian, Japanese, Spanish, and Turkish. In addition to announcing a patient assignment, the charge nurse may broadcast a message to patients in the waiting area.

2. Text processing technologies

Beyond the text search functionality found in a word processor, there are a series of text processing technologies used in AI. For example, if a social media developer wanted to know if online tweets about a company were positive or negative, he would use an AI technology named *sentiment analysis*. And if a developer at a travel company wanted to know the best travel destinations recommended by users, he would use the AI technology named *Entity Recognition* to identify cities. If a developer wants to understand the audio transcription from Siri, he has to use an AI technology named *natural language processing* or NLP.

NLP attempts to understand the meaning or intent of textual content and to do that, an NLP engine must first be able to determine in which language is the text. Next, the part of speech of the words used in the sentence, as well as other components of a language such as word stems and contractions must be identified. For example, the doctor's sentence: "I'll go to the front office" would be decomposed into the pronoun "I", the verb "will", the verb "go", the preposition "to", the determiner "the", the adjective "front", and the noun "office". So even though the doctor spoke the word "I'll" the NLP engine had to understand the concept of American English language contractions and process the two words "I" and "will". Likewise, if the

doctor declines a patient assignment by saying: "I can't go to Recovery Room 2", the word "can't" will be decomposed into the words "can" and "not".

Beyond contractions, advanced natural language processing algorithms must also consider the accent in which a user speaks. From the NLP engine's point of view, everyone has an accent whether it's the slow-taking southerner's accent, the neutral mid-Atlantic accent, the fast-taking New Yorker's accent, a Bostonian accent, or perhaps the accent of a partially deaf therapist. And of course, when we are tired, we all tend to slur our words a bit making us sound inarticulate. This also happens to surgeons and anesthesiologist after a middle-of-the-night emergency surgery. So, it should not be surprising that the intended word "I'll" can be misspoken just enough to be transcribed as the word "all". The NLP engine used in *"The Patient Is In"* apps compensates for these common transcription errors because the algorithms were tailored for a doctor's use of Siri. And as Zaid noted:

> *You should also notice that the NLP engine must be able to understand American-English Language homophones as in '8' and 'eight' and 'ate'; '4' and 'for' and 'four'; and '2' and 'to' and 'too' and 'two'.*

If we reexamine the classic NLP pipeline advocated in AI research as mentioned at the start of this section, we learned that it started with first identifying the context language and then parsed the text into parts of speech. When Zaid applied this to the doctor's English language message he saw that the phrase "front office" was decomposed and identified as the adjective "front" and the noun "office".

But, if the words in the phrase "Front Office" were capitalized, they would be decomposed and identified as the noun "Front" and the noun "Office". Because the app receives the audio transcription from the virtual assistant, which in our case is Siri, there is no way to know if the transcribed text will be the capitalized "Front Office" or the uncapitalized "front office". Similarly, there is no way to know whether the transcription will be "Operating Room 2" or "Operating Room too".

While both are linguistically correct, that does not help us because real world processes such as cleaning a room are started once the doctor has completed

assignment in a specific room. Likewise if doctor declines an assignment then another doctor will be assigned. So, the app has to ensure correctness because sometime "most likely correct" means "absolutely wrong".

Unlike the Siri messaging support in apps such as WhatsApp or Messages which simply relay the transcription provided by Siri's speech recognition, the **"Patient Is In"** must be sure which room is the doctor discussing, if the doctor is rejecting an assignment, completing an assignment, or accepting an assignment and providing an estimated time of arrival which may be provided in hours or minutes. Said Zaid:

> The NLP engine uses multiple algorithms and text processing techniques to best ensure that the doctor's intent is correctly captured even if the doctor's words were incorrectly spoken or erroneously transcribed. The approach to use multiple algorithms and extraction techniques is relatively new in AI and is called parser combinators. The NLP engine used in the **"Patient Is In"** apps on both iPhone and Apple Watch and available on the HomePod adds many Siri-specific and doctor-specific algorithms to the classic NLP algorithms.

Due to the uncertainty in processing speech, a new user interface idiom has evolved. Called the *conversational user interface*, it enables Siri to mediate a conversation between the user and the app and more abstractly between the doctor and the charge nurse.

3. Conversational User Interface

As in real life, rarely is information unambiguously clear. If you ask a taxi driver to drive you from the John Wayne Airport in Santa Ana, California to your office "on Main and MacArthur in the next town over" which is a few blocks away in the city of Irvine, he may instead drive you a few miles further to Main and MacArthur in the city of Costa Mesa as both cities are adjacent to Santa Ana. Not only do these adjacent cities have the same Main and MacArthur street names, they are same streets which intersect in two different cities. For those of you living in Washington DC you know there is a similar problem: 3232 Georgia Avenue **NW** is not 323? Georgia Avenue **NE**.

With voice interfaces, the app has to support a conversation to clarify the user's words or meaning by asking for more information from the user either because the user has not provided enough information or has provided ambiguous information. In general, the app must also confirm its understanding before taking action on behalf of the user.

Siri mediates this conversation between the user and the app. It handles the speech recognition and passes the audio transcript to the app which processes that with natural language processing algorithms and other text processing technologies. If the app needs additional information, it asks Siri to prompt the user by providing Siri with text to read to the user. Siri uses text-to-speech to ask those questions. After the app is satisfied that it understands the user's request, the app tells Siri to ask the user to confirm or reject those assumptions and with the user's permission, the app finally processes the user's request. Zaid again:

> For example, if a doctor is assigned to both "Recovery Room 1" and "Exam Room 3" and tells Siri: "I can't go", the Patient is in app will have Siri prompt the doctor to clarify if she meant "Recovery Room 1" or "Exam Room 3". If the doctor tells Siri: "I can't go to Exam Room 3", then there is no need for the disambiguation phase and Siri will only prompt the doctor to confirm that she cannot go to Exam Room 3.

4. Contextual Computing

The classical algorithms for speech recognition and natural language processing rely on a single invocation of the algorithms on a piece of audio or text content. Consequently, understanding and correctness is determined by a single utterance or phrase. Previous use of the algorithms on other content or other relevant information is not used. As Zaid pointed out:

> Systems which use those algorithms rely on the algorithms' built-in knowledge of human language and its grammar and at best may allow the user to provide a custom dictionary to identify non-standard words. This is common in advanced dictation and character recognition (OCR) software for the medical or legal

professions. iOS allows the user to provide custom pronunciations of names and that information is used in both speech recognition and speech synthesis.

Real-world AI apps and services use contextual data to augment their algorithms to make recommendations. Yet, there are some problems which even the best AI algorithms cannot solve due of the complexity of the data, quality of the data, or other perceived problems such as changes in an established pattern. When integrated with a virtual assistant, these AI apps and services can ask for assistance from the user to clarify their words or intent.

A good example of this can be seen in the ***"Patient Is In"*** app. If the NLP-engine is unable to determine the doctor's intent, Siri will prompt the doctor to disambiguate between the most likely choices the algorithms believe to be contextually relevant. So, if Siri transcribed: "I'll go to the Operating Room too" and the app was unable to determine if the doctor meant the OBGYN "Operating Room" or "Operating Room 2" in the Emergency Room, the NLP engine uses its knowledge of the active patient assignments and will have Siri prompt the doctor to choose between "Operating Room" and "Operating Room 2".

Another aspect of contextual computing, is the dynamic adaptation of content and the user interface based on the device context. For example, if a doctor was using her AirPods and received a new patient assignment, the ***"Patient Is In"*** app for iPhone would detect the presence of the AirPods and would automatically read aloud the patient assignment to the doctor rather than just generating the typical notification sound. It accomplished this by generating a textual description of the patient assignment with any notes about the patient and then used speech synthesis to pipe that audio to the AirPods. A doctor could then use Siri directly from her AirPods and respond to the new patient assignment in one fluid motion. Said Zaid:

> *Doctors loved the convenience of not having to find their phone and then access the notification to learn the details of the patient assignment. This removed stress from the doctor while reducing the patient wait time. This feature was originally designed for blind therapists who specialize in helping young people enter society*

with loss of vision and seniors who have to transition to a new phase of life without sight. This is an example where usability and accessibility are one and the same.

In addition to content and interface adaptation for AirPods, a similar process occurs if the charge nurse decides to send a patient assignment to an AirPlay compatible speaker … rather than reading out loud the patient's notes, the app will only announce the room information. The difference is one of privacy. It is acceptable for HIPAA-type information to be automatically announced over a doctor's AirPods but should not be automatically announced over a speaker even if that speaker is in the doctor's private office.

Though we have discussed contextual computing as if it is a new concept, programmers have always had to consider how their solutions will work in different environments. While web developers have an opportunity to extend a desktop browser solution to mobile users, app developers now have the opportunity to design solutions which reach into a user's physical world.

B. Ambient, Continuous, Pervasive

First seen in the sci-fi movies and literature of the 1960s and then popularized on the 1990s TV show *Star Trek: The Next Generation*; ambient intelligence, pervasive computing, and continuous computing are often described by computer researchers and futurists as the ability for a computer system to monitor and communicate both proactively and reactively to nearby people and devices. As these concepts have matured into consumer technologies and products, some companies have chosen to use the term ubiquitous computing to describe their implementation of ambient intelligence, pervasive computing, and continuous computing.

NOTE: many of you will remember the primordial real-world implementation of these concepts were the smart displays placed in each room of Bill Gates' home. Upon arriving at the former Microsoft CEO's home, each guest would wear a special badge which would identify the guest to the home's computer system. Sensors in the badge enabled the home

to track the guest and upon entering a room, the smart display in that room would cycle through a series of the guest's favorite paintings.

Adapting these concepts to a hospital setting, a doctor should receive patient assignments proactively based on her on-site location and then respond reactively by voice. ***"The Patient Is In"*** implements ambient intelligence, pervasive computing, and continuous computing to reduce patient wait times.

C. Reducing patient wait times with iBeacon & IoT

There are many benefits for the charge nurse to know the location of on-site doctors. Early in the development of the ***"Patient Is In"***, doctors identified many use cases in which they wanted the charge nurse to automatically know their location without having to manually update each charge nurse in the departments to which they report. One surgeon told Zaid of a now common problem in large hospitals: the surgeons' offices are located in an adjacent building or wing to the hospital's operating rooms and it can take up to twenty minutes to walk from his office to the surgery area. It seems that the best utilization of space is to densely co-locate surgery rooms, medical equipment, and treatment areas and that means that the commute time of personnel is a secondary priority. But for many patient assignments related to emergency surgery or consultation, it is crucial for the charge nurse to assign the closest, available doctor and not randomly select from the pool of on-site doctors for the current shift.

So, ***"The Patient Is In"*** apps use Apple's iBeacon micro-location technology to automatically register a doctor's iPhone with a charge nurse's iPad so that whenever a doctor approaches or leaves the area of a charge nurse's iPad, knowledge of the doctor's location is propagated to all charge nurses so that they can make intelligent decisions about patient assignments. Fans of Star Trek will recognize that this is a modern-day implementation of the recurring scene in which the staff query: *"Computer, locate Captain Picard"*. This feature is disabled by default and allows the doctor to choose if she wants to share her on-site location with the charge nurse.
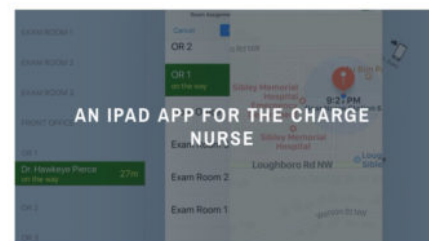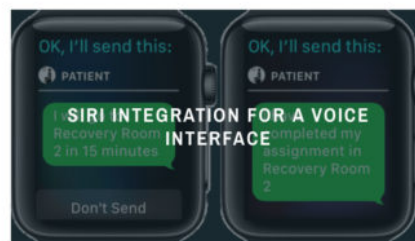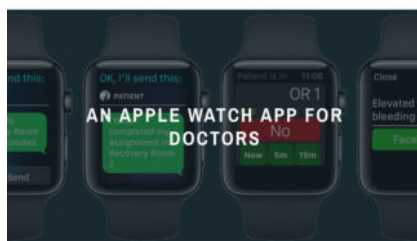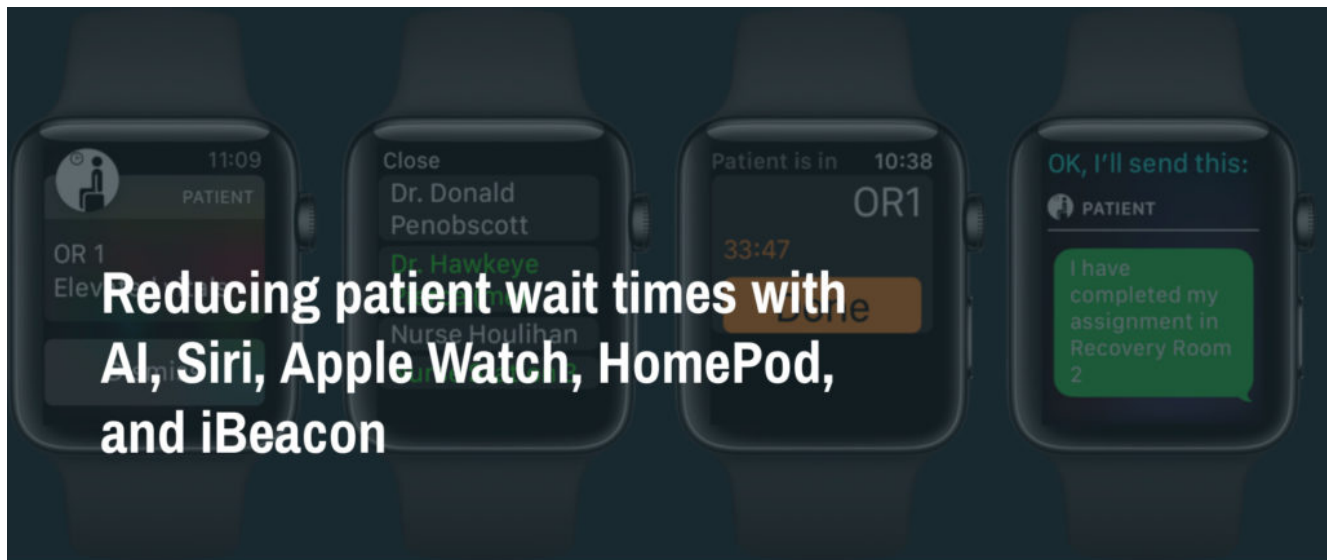
III. CONCLUSION

This was, at best, a cursory look at the use of Siri in healthcare and how it can reduce patient wait times when coupled with proactive notifications sent to a doctor's Apple Watch and wireless announcements sent to AirPlay smart speakers such as the HomePod, the latter of which deserves more space but time precluded providing more. And while voice assistants from other vendors such as Amazon and Google may perform better at web searches and commerce, Zaid has achieved something way beyond that: the integration of Siri with secure messaging, real-time notifications, wearables, proactive wireless announcements, and knowing the location of the medical staff which can significantly reducing patient wait times. His employment of ambient intelligence, pervasive computing, and continuous computing is impressive.

I'll close with a comment from Zaid:

> In the late 1990s, legendary computer scientist Bill Joy categorized how people and machines would access data over the web. He described a "Weird Web" in which a user would consume information by listening rather than reading. And, users would request information via speech rather than by typing queries on a keyboard.

*Based on my experience in developing **The Patient Is In**, **The Mashup App**, and implementing a version of ambient intelligence, pervasive computing, and continuous computing, I believe that Apple should once again reshape computing by extending Siri from a siloed, device centered service to include the concept of Siri meta-services and a virtual data store running on a group of devices.*

Follow Zaid on Twitter **@app_devl** and the app **@ThePatientIsIn**

For more information about the **Patient Is In** visit: **www.ThePatientIsIn.com**

Optional footer notes.