Siri

A Primer

Like everyone, I have been bombarded by Apple's advertising campaign announcing iPhone 4S with the Siri Digital Assistant. I spent 2 1/2 years working with interactive voice response (IVR) voice applications, mostly in the call center space. I learned enough to conclude that voice interface technology suffered from decades of over-promising and under-delivering...simply put, most people have negative experiences with voice interfaces, and little, if any, experience using a natural language interface.

Now Siri comes along and appears to have the natural language interface chops to overcome the skepticism that has built up over the decades. From what I've seen and heard, Siri appears almost magical, availability issues notwithstanding. This initial response is absolutely essential if the market is to cross chasm that was created during the more than thirty years of voice interface failures and disappointments.

The following presents a lot of research and my opinions regarding the recent release of Siri. My goal here is to consolidate all that I have gleaned into a single document that concisely yet comprehensively profiles Siri. This presents opinions and prognostications that reflects my voice technology experience and the research I've done since the 10/4/11 introduction of Siri.

Background

Siri began as a DARPA-funded project at SRI International back in 2003 and was spun out and setup as Siri International (SI) after DARPA funding ended in 2008. SI raised nearly $25 million and launched the Siri Digital Assistant as a free App on the App Store in early 2010. Four months later, Apple acquired SI for $200 - $250 million. Since then Apple integrated Siri technology with iOS and extended the backend architecture that allows Siri to work so well. The free Siri Digital Assistant App was removed from the App Store (9/15/11) and Apple ceased supporting Siri on any device other than the iPhone 4S after 10/15/11.

Approach

On high, what Apple does so well is to improve upon what others have done by making technology fit to the human experience instead of requiring humans conform to the computer experience. Siri's initial success is due to Apple's adherence to four core development principles:

1. Fairly attainable early goals, which it accomplishes well.
2. A large population of enthusiastic adopters who give it sustenance.
3. Lots of room to improve, giving Siri areas to grow into with positive feedback.
4. Apple's patient commitment and deep pockets to see Siri achieve its potential.

Services

Siri is not Hal 9000 (funny aside: after I wrote this, I read a posting about *IRIS 9000,* which was named for Hal 9000, the first Siri peripheral which allows hands free Siri interactions). The advertising campaign introducing Siri presents a totally buttoned down solution but the digital assistant really focuses on a finite number of areas. Siri works well today because Apple got Siri to do a smallish number of very meaningful things really competently. Ask Siri what "she" does and it will list these things (i.e., Siri's current "use cases"):

- Phone (to make and receive calls)
- FaceTime (to make and receive video calls)
- Music (find and play songs and playlists)
- Mail (read and write emails)
- Messages (read and write messages)
- Calendar (read and update calendar)
- Reminders (set and access reminders)
- Notes (read and write notes)
- Contacts (search contacts)
- Weather (access Yahoo weather)
- Stocks (access Yahoo finance)
- Web Search (Safari, Maps, Google)
- Find My Friends (locate other iPhones/iPads)
- Alarms/World Clock/Timer
- Wikipedia
- Wolfram|Alpha (in English only for math)

Now if you consider the list closely, what you'll notice is that it is not as open-ended as it first appears. Siri can't understand just *anything.* Siri can only do a certain set of key tasks. In a nutshell:

- Use the phone.
- Interact with the calendar.
- Search contacts (not create them).
- Read and write messages (text, SMS and email).
- Interact with the Map app and location services.
- Access certain pre-defined data providers (Yahoo! Weather, Yahoo! Finance, Yelp, Wolfram|Alpha and Wikipedia)

As impressive as this is, the above list hardly defines everything we want or need to do with mobile devices and this speaks to the Siri opportunity.

Technology

Siri is a notable implementation of several technologies: Nuance Communications' voice recognition and text-to-speech (TTS) technology, Siri's artificial intelligence-like (AI) natural language processing engine and Apple-hosted backend services (i.e., processing capabilities and access to data and other resources). Perhaps a useful simplification is to suggest that Siri has three layers: voice processing, grammar analysis-context engine and services. To these I would add user voice and crowd-sourced learning to establish Siri's technology milieu.

At the iPhone level, it appears that Apple has integrated Nuance's voice technology with iOS to allow spoken words to be translated to data (text) and to accept text and give it a voice. Nuance has been in the voice technology business since 1994 and, interestingly, was another spin out from the same lab as Siri (SRI International's STAR). It is fair to say that Siri's actual interface is really Nuance…but it's the backend magic of Siri that really makes things pop.

What Siri appears to do better than any other mobile voice solution is process natural language. This is not voice activation, where users learn to speak commands. Nor is it traditional speech recognition where enunciation is essential and only when words are said clearly enough, tasks are accomplished. Siri goes far beyond this by understanding language, modeling knowledge and applying logic.

With Siri, there are no *pre-defined* ways of requesting Siri do something or answer a question—it simply understands what a user wants to do. Importantly, Siri not only understands spoken words, it understands context. Understanding context requires deciphering natural language and then adroitly accessing the resources at Siri's disposal to perform tasks or correctly answer basic or even certain complex questions.

To add clarity to this, consider what happens when you ask Siri to, "Book a table at Beachfire in San Clemente for 5PM," Siri determines that "Beachfire" is likely the name of a place, "San Clemente" is likely a location and that Siri still needs to be sure to confirm the reservation is today. This is referred to as natural language processing, and as you can imagine, it is incredibly difficult to get right.

Siri architecture, like Android's Ice Cream Sandwich (ICS), relies upon backend processing and data access. This is largely due to mobile processor, memory and storage constraints. Viewed by most as a limitation (Siri doesn't understand "*Call home*" without network access), Apple appears to bank on the belief that network access (WiFi, WiMax, cellular, etc.) will become faster, broader and more reliable over time and that doing the heavy computational lifting natural language processing requires is best handled by powerful data center servers and their fat pipe access to resources and services. An advantage to this architecture is that extending platforms services can be done centrally and made available simultaneously to all users.

Siri also learns. At the user/handset level, Siri has routines that allow Siri to better understand the subtleties of each individual user's accent and voice characteristics. At

a macro level, Siri's backend culls through the millions of requests (think: Google Search or Apple's Genius) and finds things to improve upon. For example, when Siri first launched it voiced "Tee X" but within a week it began saying "Texas".

What really sets Siri apart, and though more of design specification versus a technology, is Siri's "friendly edginess" and humor—its persona. Siri tries very hard to be witty *and* very useful. This is very difficult but critical as this "personality" is what has captured the imagination of the market. When merited, Siri delights users with clever, cheeky and laughter-provoking responses. I very much doubt Siri would be getting all of the attention it has if Siri gave accurate but boring responses every time.

How Siri Works

Though the following needs to be confirmed, it appears that once the the Siri microphone button is touched, whatever is said is turned into text and sent to Apple's data centers where Apple hosts Siri's AI-like natural language processing engine. Siri then figures out what has been said. Depending upon the inquiry, Siri either sends a text response back to the iPhone or performs queries and sends a text response and data back to the handset. The iPhone's Siri Digital Assistant is given "life" by vocalizing answers to the question asked using the Nuance text-to-speech service and, if merited, displaying information obtained via Siri's backend services (e.g., Wolfram|Alpha data) or via the user's iPhone resources (e.g., Map app).

Importantly, Siri also somehow manages conversations—multi-part exchanges between a user's iPhone and Siri's data center resources. So when Siri needs more information to fulfill a request it asks the user for more information without forgetting what was originally asked. This is critical to making a digital assistant conversational. Siri's ability to maintain a "conversion" by supporting multiple exchanges is essential to making Siri work as smoothly as it does today.

Integration

Apple has integrated Siri with iOS5 but what does this actually mean? Turning speech into text can be done without extraordinary processing power. However, using AI-like technology to understand what has been said is generaly an intensive computational task. By integrating the means to both turn speech into text and text into speech at the handset level and running the context engine on big iron at a data center, Siri succeeds. I also think that by integrating parts of Siri with iOS, Apple is protecting how third parties can use Siri, helping to ensure that that which sets Siri apart remains a significantly Apple-only technology.

Siri was architected as an extensible platform so new services can be added without extraordinary development effort. Should Apple open Siri to developers—and I am not 100% certain that they will do this soon or at all—there is virtually no limit to the breadth of domains Siri could eventually support.

Status

Apple has designated the initial version of Siri Beta, a rarity for Apple, which is why, in part, they have not opened Siri up to the Apple development community. Currently, Apple has not announcement of when or if a Siri API/SKD will be available. Some predict Apple will release the SDK with the launch of iPad 3 in early 2012 but I believe it could be well over a year, if at all. The challenge is that Siri has a distinct persona that is controlled by one group within Apple. How can Apple open this up and maintain that efficient, compliant, engaging and uncomplaining "voice with an attitude" if thousands are developing for the platform?

I believe that Apple sees Siri's vast potential and decided to take more time to expand Siri to fully support non-US markets, improve it and extend Siri to other Apple products. For example, rumors abound that Apple will begin selling HDTVs with a Siri TV control in late 2012 or 2013 and I'd wager that Apple-TV will support Siri so that all TVs can be controlled via the Siri voice interface.

Fanboys and Apple bashers alike agree that Siri is the best implementation of voice technology to date; even "Adrodians" cede that Siri leaves anything on the Android platform in the dust. Given the potential of the technology to sell more Apple products that support Siri, it is not unreasonable to suggest that quite some time could pass before Apple feels the solution is ready to open Siri up to the general Apple dev community. But again, this may never happen.

Siri currently isn't a "ready for prime-time" Apple product which is why it is technically a Beta release. The biggest drawback is that it is not a global product. While the initial Apple version supports American and UK English, French and German, Siri's full functionality only works with American English in the US. Apple still needs to expand data centers in Europe and Asia to give folk there the full flower of the Siri experience.

Another limitation, presently at least, is that Siri operates in a closed ecosystem. It doesn't work with other apps or services other than those Apple has connected on the backend. This is just one of the challenges Apple faces in allowing others into Siri's world. Beyond technology, Apple will likely need to develop new economic models or hybrids usage/licensing and until they figure this out, developers will not be able to tap into Siri.

Apple reliance on backend servers to do a lot of the heavy computational lifting exposes two Siri other limitations: network availability (it simply doesn't work when access to the Internet is not available) and Apple data centers resources. If Siri proves wildly successful, Apple will need to scale server resources to keep pace and this is expensive and tricky. Siri's volume in the first month since release is reportedly 10 times what Apple predicted and there have already been complaints that Siri sometime just stops working. Pundits attribute at least some of these outages to Siri traffic exceeding the capacity of Apple's hosted services.

Potential

Apple's raison d'etre, natch, is to sell more Apple hardware, software and services. Apple must have concluded that Siri has the very real potential to drive many billions in new revenues otherwise their $200+ million acquisition of Siri would make little sense. Even Apple, as a public company, can't waste $200+ million of its shareholder's money without Wall Street taking notice.

Job's 2nd Apple Life was marked by the company casting an outrageously wide net—new computer hardware/devices, OS/software/apps, devices and services for music, telephony, television, books, movies, etc. Still, if you look at where Apple doesn't lead it's in areas where the default Apple approach—controlling the entire experience—cannot work due to market size, structure or other deep-pocketed, 800 lbs. gorillas. Two examples of this are iBook and Apple-TV. In both markets, Apple is not expected to dominate, just remain a player.

Apple is spectacularly successful but not by any stretch the only or even biggest game in town in most of its markets, including smartphones. Apple is where they are because they charge a premium for Apple cachet. As such, opportunities could exist to exploit what could be a Siri-driven/inspired tsunami by going where Apple won't go.

Worth noting is that Siri has the potential to be a real threat to Google. If one thinks about it, Google makes hay by presenting a bunch of answers to queries ranked from those sponsored to those "earned" in descending order of relevance according to Google's top-secret search algorithm. Now Siri comes along and offers specific answers—and from all reports, highly correct answers—without sponsorship and when merited, ranks results by objective metrics (e.g. distance from location). Siri effectively bypasses search (as we know it) entirely and this too speaks to one of a myriad radical potential changes that Siri could usher into the market.

I think that with the Beta release of Siri, Apple is working on building it for global release while figuring out what they want their slice of the overall digital assistant pie to look like. Once they do that, I believe that Apple will open Siri up in a hyper-controlled way to allow others to enrich Apple further by developing markets, apps and solutions that drive core product revenues (iPhone, iPad, apps, etc.).

Competition

Siri has very competent competition. Android's ICS Voice Actions, Voice Actions Plus, Speak With Me, Vlingo, Speaktoit Assistant, Edwin, a hack put together by an Android development team in a day called Iris (Siri spelled backwards) and doubtless many others. Speaktoit, Iris and Edwin talk back to you. Vlingo doesn't speak. However, they all make it easy to make phone calls, send messages and get information like the weather or the location of the closest cafe.

So Siri currently doesn't do anything all that different than other voice solutions that tap into users' contacts, calendars, email, etc. What is different is the experience—the persona—and this, plus Siri's ability to learn and become more and more accurate will allow for Siri to grow better and support broader and richer services. Bottom line, Siri will always have competition and the Android+voice interface will likely remain the logical alternative to iOS+voice interface.

Aside from iTunes and iPads, Apple rarely dominates markets; the Mac platform has never exceeded 20% of the total market for personal computers and iPhones will likely continue to experience erosion of its smartphone market share. It is quite safe to suggest that Siri will not be the only game in town. Competitors simply have too many ways to assail Siri. What I cannot predict is where the market will ultimately shake out...will Siri command a huge share of the market or slug it out amongst Android offerings and achieve Apple's more typical "nice" share of the premium user market. Time will tell but I predict that Siri is more iTunes/iPad like than Mac/iPhone.

Conclusions

Siri's first Apple iteration opens minds and speaks loudly to Siri's potential. What struck me is that even with this initial release one can readily imagine a sea change in the way humans interact with mobile and even stationary devices over the next several years…and this is only the beginning.

Apple has produced duds—Newton and the first Apple iPod phone (the one with Motorola), to name just two. Still, I recently encountered a non-savvy, by appearances at least, iPhone 4S owner that was literally gushing about how Siri is changing her life. Siri does not appear to be Newton II.

My voice application business allowed me to see the potential of truly great natural language voice technologies. The few great applications I found left me believing that someday, voice will handle large numbers of everyday tasks and, where appropriate, even more complex things. I think that this time has come.

In the end I believe that Siri is showing the world that a great voice interface is viable while simultaneously giving everyone amnesia about past failures. For the first time we have a "good enough" general-purpose and natural language-based voice interface technology, which means that there is no going back. Most important, to me at least, Siri opens the door to huge opportunities to catch the Siri/ICS wave and profitably exploit these technologies in new and exciting ways.

Select Research Links

http://www.jeffwofford.com/?p=817 (Must Read: Concise Siri summary)

http://learninglayer.wordpress.com/2011/11/06/siri-an-it-inflection-point/ (Must Read: The importance of Siri)

http://labs.vectorform.com/2011/10/the-impact-of-apple's-siri-release-from-the-former-lead-iphone-developer-of-siri/ (Must Read: Siri developer perspective)

http://techcrunch.com/2011/11/09/gary-morgenthaler-siri-will-eat-google/ (Must Read: Siri potential/Google threat)

http://www.pcworld.com/article/242198/apples_siri_vs_androids_voice_actions_feature_showdown.html (Apple Siri versus Androids Voice Actions)

http://thenextweb.com/media/2010/08/22/kindle-outselling-ibooks-60-to-1/ (iBooks versus Kindle)

http://www.asymco.com/2011/10/18/clayton-christensen-and-siri/ (Historical perspective)

http://www.electronista.com/articles/11/07/22/ios.5.to.integrate.siri.with.crowd.sourcing/ (early announcement of Siri iOS integration)

http://appadvice.com/appnn/2011/08/apples-sdk-sheds-some-more-light-on-ios-5-siri-integration (Siri SDK)

http://androidcommunity.com/ios-5-functionality-has-already-existed-on-android-20111014/ (Siri versus Android)

http://www.pcmag.com/article2/0,2817,2394787,00.asp#fbid=OO9YCYpwcTb (General information)

https://discussions.apple.com/thread/3388900?start=105&tstart=0 (Siri problems)

http://blogs.wsj.com/tech-europe/2011/10/26/siri-attracts-competition-and-its-first-peripheral/ (Competitive offerings)

http://tech2.in.com/news/hdtvs/apple-sirienabled-hdtv-has-been-in-the-works-for-years/252752 (Rumor that Apple will release a Siri-enabled TV)

http://shitsirisays.com/ (Siri is edgy/funny)

http://www.oulixeus.com/2011/10/want-to-know-what-the-next-generation-of-apps-will-be-like-talk-to-siri/ (Siri SDK)

http://en.wikipedia.org/wiki/Nuance_Communications (Nuance)

http://forums.macrumors.com/showthread.php?t=1268497 (Does Siri learn)

http://www.apple.com/iphone/features/siri-faq.html (Siri learning)