Getting Rid of Wake Words: Moving Towards Seamless Conversations with Intelligent Personal Assistants

Shashank Ahire
Leibniz University Hannover
Germany, Hannover, 30167
shashank.ahire@hci.uni-hannover.de

CHI'20,, April 25–30, 2020, Honolulu, HI, USA
ACM 978-1-4503-6819-3/20/04.
https://doi.org/10.1145/3334480.XXXXXXX

Abstract
In this position paper, we aim to draw the attention towards the wake words. The wake words are an integral part of every request addressed to the IPA(Intelligent Personal Assistants). Currently, we use a wake words to initiate the conversation with IPA. Every request made to the IPA is led by the wake words, making a conversation with an IPA more tiresome than a conversation with a human being. It seems that there is lack of discussion of whether wake words are essential or whether there are better alternatives. Can we get rid of wake words completely or at least in specific contexts? In this position paper we highlight the role of wake words and the problems they create. Based on our experience with IPAs we propose three less burdensome alternatives that avoid the problems of wake words in some cases. The proposed approaches are inspired by conversation events in our daily life. From these approaches we aim for a seamless conversation with IPAs.

Author Keywords
Intelligent Personal Assistant(IPA); voice assistant; Conversational User Interface(CUI); wake words

CCS Concepts
•Human-centered computing → Human computer interaction (HCI);

Introduction

wake words

In the our day-to-day interaction with IPA we use wake words. Every IPA (Intelligent personal Assistant) has a wake words. The wake words is generally, used to initialize the IPA and listen to the user request. The formal definition of wake words according to the Cambridge dictionary is "word or words that you say in order to make an electronic device, or a feature on a device, ready to work"[6]. Many of us have already used wake words in our daily life and we know the important role it plays when interacting with a voice assistant. Today, all IPA we know of use a sequence of one or two words as wake words. They are typically words that are easy to remember and articulate. Often, the wake words is the name of the company or application.

Purpose of wake words

Considering the basic functions of the IPA, of performing a search and executing a task requested by user. To perform these functions, it is necessary for the IPA to know which task or questions are addressed to it. For identifying, the request and questions it needs some cue from the user during the request. Also from our day-to-day experience with IPA, we know wake words play an important role in the request and without it the IPA won't be able to detect our request.

To capture the cue in the form of wake words, the device is constantly in the hearing mode. After capturing the wake words the device initiates the process of listening and recording sound until it detects a few milliseconds of silence. The silence serves as the second cue, indicating the end of the request. As soon as the IPA detects the end of the request, it sends the recorded audio to the cloud for interpretation. Based on the interpretation, a response is sent to the device from the cloud and is given to the user[7]. In the absence of wake words the system would not be able to detect if the utterance was directed to the IPA.

If the IPA was designed without wake words, then it would have to always keep itself in listening mode and keep interpreting and responding to each and every potential request it hears. This approach would create a substantial amount of unnecessary processing followed by the IPA response, eventually leading to audio clutter. Additionally it would also waste resources like network bandwidth and electricity. But most importantly, since IPAs are many times placed in sensitive settings, like the home or the office, an always-on listening mode would breach the privacy of the user.

Considering the above reasons it perfectly makes sense to have wake words for interacting with IPAs. Hence, the device can go to sleep mode or low power mode, and wake up on audio signals and just try to recognize the wake words. Only after detecting the wake words it would continue the process, thus saving time and effort. In this approach, considering factors like privacy, efficiency, and speed, wake words serve the purpose suitably.

Apart from the technical purpose of serving cue, users are using them for different functions such as to stop the IPA when it gives a wrong response, when IPA interrupts the user before completing the request or in the situation when the IPA answers are too long and irrelevant. Also, they used the wake words to take the control of the conversation[9].

Alternate wake words

The majority of the IPA has a fixed wake words. As mentioned in Table 1 and from our personal experience with the IPA, we have come across some wake words that are easier to pronounce and remember whereas some are tough. Many of users have reported this issue in the various online forums of the respective companies [2, 1, 4]. Due to incon-

IPA	*wake words*
Google Assistant/Nest Hub	Ok Google
Amazon Alexa(Echo/Dots)	Alexa, Amazon, Computer
Apple's Siri	Hey siri
Microsoft's Cortana	Hey Cortana
Samsung Bixby	Hey Bixby

Table 1: IPAs and their respective wake words.

venience some users have found their own hack for the wake words which are easier and funnier than the original. Especially for the wake words "Ok Google" users are using wake words like "Boo boo," "Hey Doodle," "Egg Noodle," etc. [13]

Currently, Amazon Alexa provides three alternative wake words. The user can choose any one of the wake words mentioned in Table 1. Since, these devices are shared among family members, it is mandatory for all family members to agree on one wake words. Among all the IPAs only Microsoft Cortana allows the user to add their own wake words [8].

Both Google Home and Alexa IPA offer a followup option, which keeps the device awake for a number of seconds after hearing a request. This technique provides some relief for the user, as they do not have to say the wake words again at least for next few immediate requests. Unfortunately, this technique only applies if the followup request is made within a few seconds. This tight time constraint is supposedly keeping the technique from becoming a great success.

Problems with wake words

The first obvious problem with wake words is its mandatory placement in every sentence. Further, it has to be placed before the request. However, sometimes we are accustomed to place the name at the end of the sentence. Eventually, it creates a burden on the user to frame the sentence in a way that the IPA gets the required cue. Thus, IPAs need to learn to extract the cue from the different types of sentence structures, which users are accustomed to use.

In probably all cultures there is a tradition of assigning pet names or nicknames for family members. Such names are generally a short or melodic. But, in case of IPAs there is no provision of assigning a pet name or short name to call the IPA. Even the shortening the current wake words does not work for example just saying "Google" will not do the job and the user has to utter it completely. Same problem is applicable to "Hey Siri" , "Hey Cortana"and "Hey Bixby."

In the general human-human communication, we have a freedom to place the name of the person before or after the request. Also, in certain cases placing the name can be omitted. For instance, consider the following scenario in which only Ben and his mom are currently in the house.

1. Mom: "Could you please turn the radio on?"
2. Ben: "OK, mom."
3. Mom: (after turning on the radio) "Thanks."
4. Mom: (after 10 minutes) "Could you please change the station?"

In this scenario, the name is not mentioned in any sentence. In the first and fourth utterance Ben understands that his mom is addressing the request to him. Furthermore, in

the fourth utterance, his mom does not have to define the context again and it is understood by the Ben that she is talking about the radio. Also, while interacting with an IPA in many cases user should not have to restate the context again and again. In the conversation above, the sentence structure in the first request is not grammatically correct, but it is clearly understood by Ben. Unfortunately, while dealing with an IPA, the it is mandatory for the user to follow the sentence structure of name first followed by the request.

The users have reported about the “mouthfulness” and “creepiness” associated with wake words, particularly “OK Google”[1][12] . Additionally, [12] also discusses wake words and the robotic nature of the wake words. Further, constantly repeating the name of the company reminded people that they are sharing their information with a company that had issues with privacy in the past. Also, Jung et al. reported some user might connotate some negative feeling with the name of IPA. They related the smartness of IPA with number of times they had to use the wake words for interacting with it. Lastly, they wanted to customise the wake words and use multiple wake words assigned for various purposes. [9].

Even though IPAs only have to recognise a particular wake words, some IPAs perform this task poorly. Sometimes, an IPA is initiated by similar-sounding words. As reported in the Amazon forum [4], a user was annoyed when Alexa awoke every time he called his daughter Alyssa. Similar problems occurred when this user had a conversation with his friend Alexis [10]. The same unintended wakeup happened with Siri and a user called Sarah [2].

The problems with the wake words are not only technical but now it has turned into a social. People sharing the name of the IPA’s particularly with Amazon’s Alexa are now potential target of bullying and awful jokes [5, 3]. This name sharing is leading to suffering and frustration for many, in some cases the children’s are victim of these uncustomisable names. Due this annoyance many of them are changing their names[11].

Proposed Solution

Dedicated Mode for Interaction
Often IPAs are placed in rooms such as halls or kitchens. Since these rooms are used by multiple family members, the device as well is shared by the family members. If a single family member has a longer conversation with the device for an extended period of time, the device could enter a dedicated mode, maybe also based on characteristics of the user’s voice.

During a dedicated conversation, the device listens to requests of the user explicitly. In this mode, the device could have a provision of ignoring the wake words and simply responding to any request by the user. In the this mode, the conversation is similar to a telephone conversation with another person. During telephone conversations, we are only communicating to the person on the call. Every sentence said by the one participant is assumed to be addressed to the other participant.

With this mode a problem arises, when a third person needs the urgent attention of the participant, and the participant switches the conversation to the third person, keeping the participant on the phone on hold. While interacting with IPA, this holding up and switching the conversation could be difficult.

Context-Specific Request Seeking
Most often in some context, users are likely to opt for certain commands only. In such a contexts, it is very unlikely that the user will request for something completely different or out of context. For instance, consider the scenario in

which the IPA is assisting the user during cooking. In this scenario, the user is likely to request instructions that are related to the current activity of cooking. It is very likely that the user will request for the next ingredient, the previous ingredient, the next step, the previous step, etc. Similarly, in case of a music app, the user is likely to ask for playing another song, stopping, pausing, changing the song to a specific song, etc.

The context-specific conversation approach often occurs in human-human conversation. In this case the context of the conversation is already established and both the participants are aware of the context. Based on the established context, the further conversation is built up. Moreover, after establishing context the participants do not address each other by name. The example given above of Ben and his mom conversation shows this reliance on context.

To make the conversation between IPA and human more *'humanlike'*. It is compulsory, to remove the need for the wake words, when the context is already set. After setting up the context, the IPA would be seeking for the most likely requests related to the current context. For doing so, IPA could cache the requests used in the current context to establish a model of probabilities of certain utterances in given contexts. If the request is related to the ongoing context it would then not be mandatory for the user to add the wake words. Also, based on the voice recognition engine in the IPA, it would be easier for IPA to identify the user and cache the next likely instruction depending on the previous session with that user.

If the user wants to switch to another context, the IPA could switch to the requested context by recognizing the user's voice. However, if the another user talks to the IPA for a different task, the IPA would have to be started with the wake words.

Voice Recognition and Learning

Voice recognition is an underutilized feature of IPAs. Using this feature the IPA could track the request of the user based on their voice and speculate the upcoming instruction in a similar way as with context-specific requests discussed above. This approach is very similar to the approach human's have in a conversation. First the IPA recognizes the user's voice, as we do in human-human conversations. Next, based on the recent interaction the IPA setup the context and communicates based on the information available. This technique would give a more "humanlike" feeling to the conversation.

Conclusion

As IPAs are becoming ubiquitous, it is necessary to the make the interaction with them seamless. The wake words are important for having a conversation with IPAs, as they serve a cue. However, in some contexts the wake words is repetitive and does not have much significance. In order to make interaction with IPA more "humanlike" it mandatory for the researchers in conversational user interface to find a strategy to reduce the unnecessary repetition of wake words and make them customisable. In this position paper we tried to highlight the role of wake words, why they are so important, what purpose they serve in our conversation, what could be possible strategies to overcome them. We presented three solutions inspired from our day-to-day conversation events to overcome the repetitiveness of the wake words. These are dedicated interaction modes, context-sepecific request seeking, and voice recognition and learning. Certainly, developing these approaches in detail and finding other ways to overcome the need for wake words are required next steps.

REFERENCES

[1] 2018a. (1st world problem) "OK Google" is too much of a mouthful. (2018). https://www.reddit.com/r/googlehome/comments/69ibyy/1st_world_problem_ok_google_is_too_much_of_a/

[2] 2018. Change Siri's name. (March 2018). https://discussions.apple.com/thread/8319320

[3] 2018b. People named Alexa. How has your life changed? (2018). https://www.reddit.com/r/AskReddit/comments/7rc7yq/people_named_alexa_how_has_your_life_changed/

[4] 2019. custom wake words. (Feb 2019). https://www.amazonforum.com/s/question/0D54P00006zSt1JSAS/custom-wake-words

[5] 2019. From all the people named "Alexa" out there... Fuck you Amazon Echo! (2019). https://www.reddit.com/r/amazonecho/comments/98hufr/from_all_the_people_named_alexa_out_there_fuck/

[6] Cambridge Dictionary. 2020. Wake word. (2020). Retrieved February 6, 2020 from https://dictionary.cambridge.org/us/dictionary/english/wake-word

[7] Josh Hendrickson. 2019. How Alexa Listens for Wake Words. (12 July 2019). Retrieved February 6, 2020 from https://www.howtogeek.com/427686/how-alexa-listens-for-wake-words/

[8] IEvangelist. 2019. Create custom keywords - Speech service - Azure Cognitive Services. (Nov 2019). https://docs.microsoft.com/en-us/azure/cognitive-services/speech-service/speech-devices-sdk-create-kws

[9] Hyunhoon Jung and Hyeji Kim. 2019. Finding Contextual Meaning of the Wake Word. In *Proceedings of the 1st International Conference on Conversational User Interfaces (CUI '19)*. Association for Computing Machinery, New York, NY, USA, Article Article 5, 3 pages. DOI: http://dx.doi.org/10.1145/3342775.3342805

[10] Kieren McCarthy. 2016. OK Google, Alexa, why can't I choose my own safe, er, wake word? (Oct 2016). https://www.theregister.co.uk/2016/10/05/google_why_cant_we_choose_wake_words/

[11] Annie O'Sullivan. 2018. Distraught Mom Pens Letter to Amazon CEO Claiming Daughter Named Alexa Is Constantly Bullied. (Dec 2018). https://www.nbcnewyork.com/news/local/distraught-family-pens-letter-to-amazon-over-bullying-issues-with-1816730/

[12] Dave Smith. 2016. There's one big problem with Google Assistant: Saying 'OK Google' is super creepy. (Oct 2016). https://www.businessinsider.com/google-assistant-ok-google-is-super-creepy-2016-10?r=DE&IR=T

[13] Damien Wilde. 2020. Some of the best alternative 'Hey Google' hotwords for Google Home. (Jan 2020). https://9to5google-com.cdn.ampproject.org/v/s/9to5google.com/2020/01/14/weird-hey-google-alternatives/amp/?usqp=mq331AQCKAE%3D&amp_js_v=0.1#referrer=https%3A%2F%2Fwww.google.com&amp_tf=From%20%251%24s&ampshare=https%3A%2F%2F9to5google.com%2F2020%2F01%2F14%2Fweird-hey-google-alternatives%2F