



# Tired of Wake Words? – Moving Towards Seamless Conversations with Intelligent Personal Assistants

**Shashank Ahire**

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

**Michael Rohs**

Leibniz University Hannover  
Hannover, Germany  
michael.rohs@hci.uni-hannover.de

## ABSTRACT

In this paper, we aim to draw attention towards wake words. Wake words are an integral part of every request addressed to Intelligent Personal Assistants (IPAs). Currently, a request made to an IPA is led by wake words, making a conversation with an IPA more tiresome than a conversation with a human being. The main question we pose in this paper is, whether we can eliminate the use of wake words at least in specific contexts. Based on our experience with IPAs we propose three less burdensome alternatives that avoid the need for speaking wake words in some cases. Based on these approaches we discuss how to design seamless conversations with IPAs.

## CCS CONCEPTS

• **Human-centered computing** → **Personal digital assistants.**

## KEYWORDS

Speech interface, conversational user interface, voice assistant, wake words

### ACM Reference Format:

Shashank Ahire and Michael Rohs. 2020. Tired of Wake Words? – Moving Towards Seamless Conversations with Intelligent Personal Assistants. In *2nd Conference on Conversational User Interfaces (CUI '20)*, July 22–24, 2020, Bilbao, Spain. ACM, New York, NY, USA, 3 pages. <https://doi.org/10.1145/3405755.3406141>

## 1 INTRODUCTION

The basic function of an Intelligent Personal Assistant (IPA) is to perform a search or to execute a task requested by the user. Currently every IPA has a wake word or set of wake words available for the user to select from. Wake words are generally used to initialize the IPA to listen to the subsequent user request.

Current IPAs use a sequence of one or two words as wake words. They are typically words that are easy to remember and articulate. Often, wake words contain the name of the company or application. To execute a particular request the IPA needs some measure to identify requests addressed to it. For determining the user's requests, they are dependent on cues from the user. In every request, these

wake words serve as the required cue. In the absence of wake words the system would not be able to detect if the utterance was directed towards the IPA.

To capture wake words, the IPA is constantly listening. After recognizing the wake words the IPA 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 IPA from the cloud and is given to the user [9].

Another alternative for wake words offered by current IPAs are follow-up options. The follow-up mode keeps the device attentive for a number of seconds after making a request. This technique provides some relief to the user, but, technique is only helpful if the follow-up request is made within a few seconds. Unfortunately, this tight time constraint is supposedly keeping the technique from becoming more broadly used.

Although, in prior research IPAs have been longitudinally studied with diverse users and in different contexts [6–8, 11, 15, 16], to the best of our knowledge, issues related to wake words have not been emphasized before.

## 2 USABILITY PROBLEMS OF WAKE WORDS

The first obvious problem with wake words is their mandatory placement at the beginning of every utterance directed towards the IPA. However, sometimes we are accustomed of placing the name of the addressee at the end of the sentence. Thus, while interacting with the IPA, it creates a burden on the user to frame each sentence in such a way that the IPA gets the cue [8, 11].

Due to the inconvenience of uttering long wake words some users have found their own hack and formed wake words that are easier and funnier than the original. Particularly, for the wake word “Ok Google” users have come up with combinations like “Boo boo”, and “Egg Noodle,” etc. [18]. Users have reported about the “mouthfulness” and “creepiness” associated with wake words [1][17]. Also, some users do not like the robotic nature of wake words[17]. Among all the IPAs currently only Amazon Alexa provides three options for selecting wake words and Microsoft Cortana allows the user to add their own personalized wake words [10].

In the general human-human communication, we have the freedom to place the name of the addressed person either before or after the request. In certain cases, the name can even 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.”

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

*CUI '20*, July 22–24, 2020, Bilbao, Spain

© 2020 Association for Computing Machinery.

ACM ISBN 978-1-4503-7544-3/20/07...\$15.00

<https://doi.org/10.1145/3405755.3406141>

- (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 request. 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 Ben that she is talking about the radio. Similarly, while interacting with an IPA, in many cases the user should not be required to restate the context again and again. Also, in the conversation above, the sentence structure in the first request is grammatically incorrect, but it is clearly understood by Ben. Unfortunately, while dealing with an IPA it is mandatory for the user to follow the grammar rules, start with a wake words and followed by the actual request.

In probably all cultures there is a tradition of assigning pet names or nicknames to family members. Such names are generally short or melodic. For instance, mom instead of mother in the above example. But, in the case of IPAs there is no provision for assigning a short name. Simply abbreviating the predefined wake words, e.g., just saying “Google” is not sufficient to trigger the IPA; the user has to utter the sequence completely. The same problem applies to the sequences “Hey Siri,” “Hey Cortana,” and “Hey Bixby.”

Even though IPAs only have to recognize fixed wake words, some IPAs perform this task poorly. Sometimes, an IPA is also triggered by similar-sounding words or names. As reported in articles and forums, IPAs are unintentionally waking up when they hear words or names identical to wake words [2, 4, 12]. Further, name clashes lead to suffering and frustration for many. In some cases children are victims of these uncustomizable wake words. Due to this annoyance some are even changing their names [14]. Also, some people are now potential targets of bullying and awful jokes [3, 5].

### 3 PROPOSED SOLUTION

Inspired from our approach in daily conversations, we would like to propose three solutions. These solutions are based on the duration, the context and learning from previous conversations.

#### 3.1 Dedicated Mode for Interaction

During telephone conversations, every sentence said by one communication partner is assumed to be addressed to the other communication partner. If the user is having a longer conversation with the IPA, the user can verbally request the IPA to switch to a dedicated mode of conversation.

In this mode, the device listens to user’s requests explicitly. The IPA could have a provision of directly responding to the user request. This form of conversation is similar to having a telephone conversation with another person. If the IPA does not receive any request after a particular time interval, it switches back to normal mode.

In this mode a problem arises, when a third person needs an immediate attention of one of the person in conversation. If the person suddenly switches the conversation to the third person, keeping the IPA “on hold”. This scenario of holding up and switching the conversation could be difficult to implement reliably.

#### 3.2 Context-Specific Request Seeking

Most often in some context, users are likely to opt for certain commands only. In such contexts, it is very unlikely that the user will request for something completely distinct or out of context. For instance, consider the context in which the IPA is assisting the user for cooking. In this context, the user is likely to request instructions that are related to the current activity of cooking. Thus, the probability of commands like the next step, the previous step, etc. is higher.

Context-specific conversations very often occur in human-human conversation. Based on the established context, the further conversation is built up. Moreover, after establishing the context the participants do not continue to address each other by name. The example discussed above, Ben and his mom engaging in a short conversation shows this reliance on context.

To make the conversation between IPA and human more “humanlike,” after setting up the context, the IPA would be seeking for the most likely requests related to the current context. For doing so, the IPA could cache the requests used in the current context to establish a model of probabilities of certain utterances in the given context. If the request is related to the ongoing context it would then not be mandatory for the user to add wake words.

#### 3.3 User Recognition and Learning

User recognition, i.e., identifying the user by relying on properties of the voice, is an underutilized function in IPAs. By recognizing the voice, the IPA could keep track of who is making an utterance. For performing this, first the IPA recognizes the user’s voice, next it speculates the upcoming request by learning from the previous requests of the user. This technique would give a “humanlike” impression to the conversation. In this mode the IPA will seek for a specific set of commands considering the current state. The IPA will execute a request, even if the user says that particular command only.

### 4 CONCLUSION

An important issue that has to be addressed in seamless conversations is how privacy is ensured. Currently, wake word detection is performed locally and the subsequent recognition and execution is performed online [13]. If the system continuously recognize utterances of the user then we need to also ensure the privacy of the users. One way of ensuring privacy is to enable the IPA to recognize and execute some commands locally.

In order to initiate any of the above solutions, the user has to explicitly select a particular mode. This change of mode, is specifically notified to the user through a change of a voice tone of an IPA or through the change color in graphical interface.

As IPAs are becoming ubiquitous, it is necessary to make the interaction with them seamless. To make interaction with IPAs “humanlike” it is mandatory to reduce the use of wake words and make them customizable. In this position paper we discussed the importance of wake words, usability problems with them and solutions we can adopt to overcome it. Certainly, developing and evaluating these approaches in detail is future work.

## REFERENCES

- [1] 2018. (1st world problem) "OK Google" is too much of a mouthful. [https://www.reddit.com/r/googlehome/comments/69ibyy/1st\\_world\\_problem\\_ok\\_google\\_is\\_too\\_much\\_of\\_a/](https://www.reddit.com/r/googlehome/comments/69ibyy/1st_world_problem_ok_google_is_too_much_of_a/)
- [2] 2018. Change Siri's name. <https://discussions.apple.com/thread/8319320>
- [3] 2018. People named Alexa. How has your life changed? [https://www.reddit.com/r/AskReddit/comments/7rc7yq/people\\_named\\_alexahow\\_has\\_your\\_life\\_changed/](https://www.reddit.com/r/AskReddit/comments/7rc7yq/people_named_alexahow_has_your_life_changed/)
- [4] 2019. custom wake words. <https://www.amazonforum.com/s/question/0D54P00006zSt1JSAS/custom-wake-words>
- [5] 2019. From all the people named "Alexa" out there... Fuck you Amazon Echo! [https://www.reddit.com/r/amazonecho/comments/98hufv/from\\_all\\_the\\_people\\_named\\_alex\\_out\\_there\\_fuck/](https://www.reddit.com/r/amazonecho/comments/98hufv/from_all_the_people_named_alex_out_there_fuck/)
- [6] Frank Bentley, Chris Luvogt, Max Silverman, Rushani Wirasinghe, Brooke White, and Danielle Lottridge. 2018. Understanding the Long-Term Use of Smart Speaker Assistants. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 3, Article Article 91 (Sept. 2018), 24 pages. <https://doi.org/10.1145/3264901>
- [7] Leigh Clark, Nadia Pantidi, Orla Cooney, Philip Doyle, Diego Garaialde, Justin Edwards, Brendan Spillane, Emer Gilmartin, Christine Murad, Cosmin Munteanu, Vincent Wade, and Benjamin R. Cowan. 2019. What Makes a Good Conversation? Challenges in Designing Truly Conversational Agents. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, Article Paper 475, 12 pages. <https://doi.org/10.1145/3290605.3300705>
- [8] Benjamin R. Cowan, Nadia Pantidi, David Coyle, Kellie Morrissey, Peter Clarke, Sara Al-Shehri, David Earley, and Natasha Bandeira. 2017. "What Can i Help You with?": Infrequent Users' Experiences of Intelligent Personal Assistants. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '17)*. Association for Computing Machinery, New York, NY, USA, Article Article 43, 12 pages. <https://doi.org/10.1145/3098279.3098539>
- [9] Josh Hendrickson. 2019. How Alexa Listens for Wake Words. Retrieved February 6, 2020 from <https://www.howtogeek.com/427686/how-alex-listens-for-wake-words/>
- [10] IEvangelist. 2019. Create custom keywords - Speech service - Azure Cognitive Services. <https://docs.microsoft.com/en-us/azure/cognitive-services/speech-service/speech-devices-sdk-create-kws>
- [11] Ewa Luger and Abigail Sellen. 2016. "Like Having a Really Bad PA": The Gulf between User Expectation and Experience of Conversational Agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. Association for Computing Machinery, New York, NY, USA, 5286–5297. <https://doi.org/10.1145/2858036.2858288>
- [12] Kieren McCarthy. 2016. OK Google, Alexa, why can't I choose my own safe, er, wake word? [https://www.theregister.co.uk/2016/10/05/google\\_why\\_cant\\_we\\_choose\\_wake\\_words/](https://www.theregister.co.uk/2016/10/05/google_why_cant_we_choose_wake_words/)
- [13] Tim Moynihan. 2017. Alexa and Google Home Record What You Say. But What Happens to That Data? <https://www.wired.com/2016/12/alex-and-google-record-your-voice/>
- [14] Annie O'Sullivan. 2018. Distraught Mom Pens Letter to Amazon CEO Claiming Daughter Named Alexa Is Constantly Bullied. <https://www.nbcnewyork.com/news/local/distraught-family-pens-letter-to-amazon-over-bullying-issues-with-daughter-who-shares-name-with-alex-a-device-jeff-bezos-ceo-massachusetts/1816730/>
- [15] Jennifer Pearson, Simon Robinson, Thomas Reitmaier, Matt Jones, Shashank Ahire, Anirudha Joshi, Deepak Sahoo, Nimish Maravi, and Bhakti Bhikne. 2019. StreetWise: Smart Speakers vs Human Help in Public Slum Settings. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, Article Paper 96, 13 pages. <https://doi.org/10.1145/3290605.3300326>
- [16] Simon Robinson, Jennifer Pearson, Shashank Ahire, Rini Ahirwar, Bhakti Bhikne, Nimish Maravi, and Matt Jones. 2018. Revisiting "Hole in the Wall" Computing: Private Smart Speakers and Public Slum Settings. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, Article Paper 498, 11 pages. <https://doi.org/10.1145/3173574.3174072>
- [17] Dave Smith. 2016. There's one big problem with Google Assistant: Saying 'OK Google' is super creepy. <https://www.businessinsider.com/google-assistant-ok-google-is-super-creepy-2016-10?r=DE&IR=T>
- [18] Damien Wilde. 2020. Some of the best alternative 'Hey Google' hotwords for Google Home. [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](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)