



## Voice navigated apps

### How can you start using it?



**Robert van Loghem**

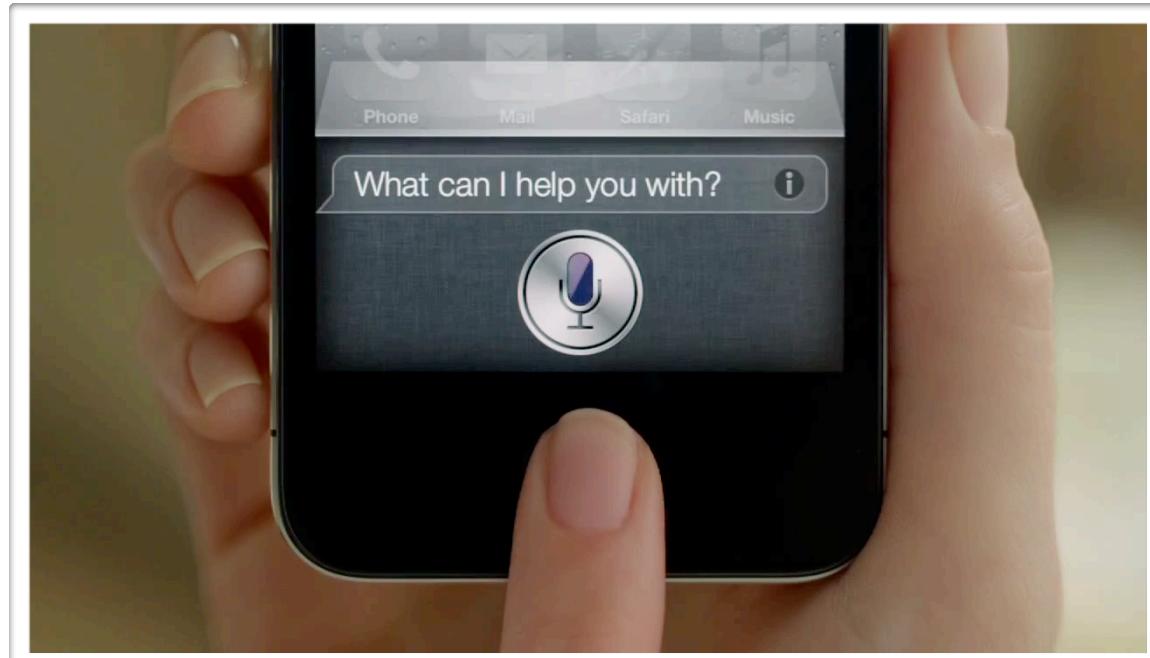
iOS developer  
@soundbites  
[rvanloghem@xebia.com](mailto:rvanloghem@xebia.com)





1985





2011



Why is  
**voice**  
interesting for  
apps?

What about visually impaired people?



It's a **faster**  
way for  
complicated  
user *input*



Banking - “*Transfer 300 Euros from my savings account to my wife’s private account*”



Home heating app - “*I’ll be home 2 hours early so set the temperature to 19 degrees...*”

Speech to speech translation - “*How much does that cost?*” - “*Hoeveel kost dat?*”

## Good Examples

***Remember!***  
***Context is a killer feature!***

Another  
example....;(



since ~2010's





TV - “*Volume up*”

Newspaper - “*Next page*”

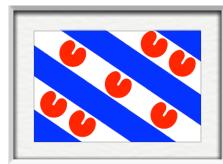
Device - “*Turn bluetooth on*” :-)

Less... good  
**Examples**



Speech  
recognition  
(ASR) has  
come a long  
way...

And also the hardware



Ljouwert,  
Liwwadden



Leeuwarden  
Maastricht



Mestreech

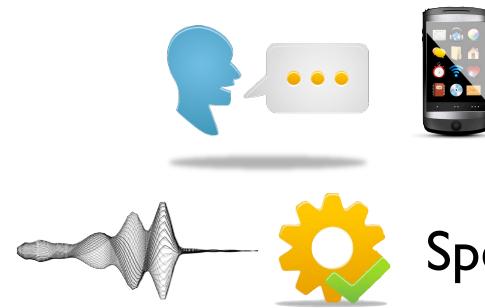


How did we get  
the voice in our  
app up and  
**running...**

(POC in **6** hours)



## Amazingly easy to unit test



Speech to text (ASR) Nuance, Google

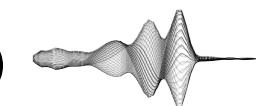
“Wanneer gaat de trein naar Amsterdam als ik aanstaande maandag om half zeven wil aankomen”



Analyze text, fetch & show results and generate text to talk back



Text to speech (TTS)



Didn't use fancy  
NLP, just find  
keywords

Using voice takes  
**3 seconds**  
to issue command



Using touch,  
it takes about  
**12 seconds**  
and *10 clicks*





What did we learn  
when going from  
**POC** to  
**production** and  
getting  
**users feedback?**





Add support for  
**hard to do  
common tasks**



**Determine vocabulary**  
and make sure it  
uses words used  
in your app



*“Ik wil naar station Woerden”*



*“Ik wil naar de \*\*\*\*\*”*

*Translated;*

*“I want to go to the red light district”*

Warning!  
test the **ASR**  
engine!



Make sure  
your users get  
**help**



The collage includes the following screens:

- Spraakherkenning**: A yellow screen asking "Ik wil morgen van Amsterdam naar Utrecht" with a microphone icon and "Klik om te spreken".
- Reisplanner**: A search screen showing "Van: Utrecht Centraal", "Via: Station", "Naar: Maastricht", and a "Plannen" button.
- Kies een reis**: A list of travel options for "maandag 24 september" from Utrecht Centraal to Maastricht at 06:44, including:
  - 03:38 → 05:34 0 1:56 >
  - 04:08 → 06:04 0 1:56 >
  - 04:38 → 06:34 0 1:56 >
  - 05:08 → 07:04 0 1:56 >
  - 05:38 → 07:34 0 1:56 >
- Reisadvies**: A summary of the selected trip: "maandag 24 september Spoor", "Utrecht Centraal 12", "Maastricht 3", with departure at 04:38 and arrival at 06:34.
- Details**: A breakdown of the trip: "04:38 Utrecht Centraal 12 Intercity" and "06:34 Maastricht 3".
- Prijs voor deze reis**: Pricing information: "Kortingspercentage Vol tarief 20% 40%", "2e klas".

# Don't mimic your normal UI flow



5% of our  
users, use the  
beta  
functionality  
**every day**

but manage their expectations  
(ASR gets better with more use),  
explain the magic!

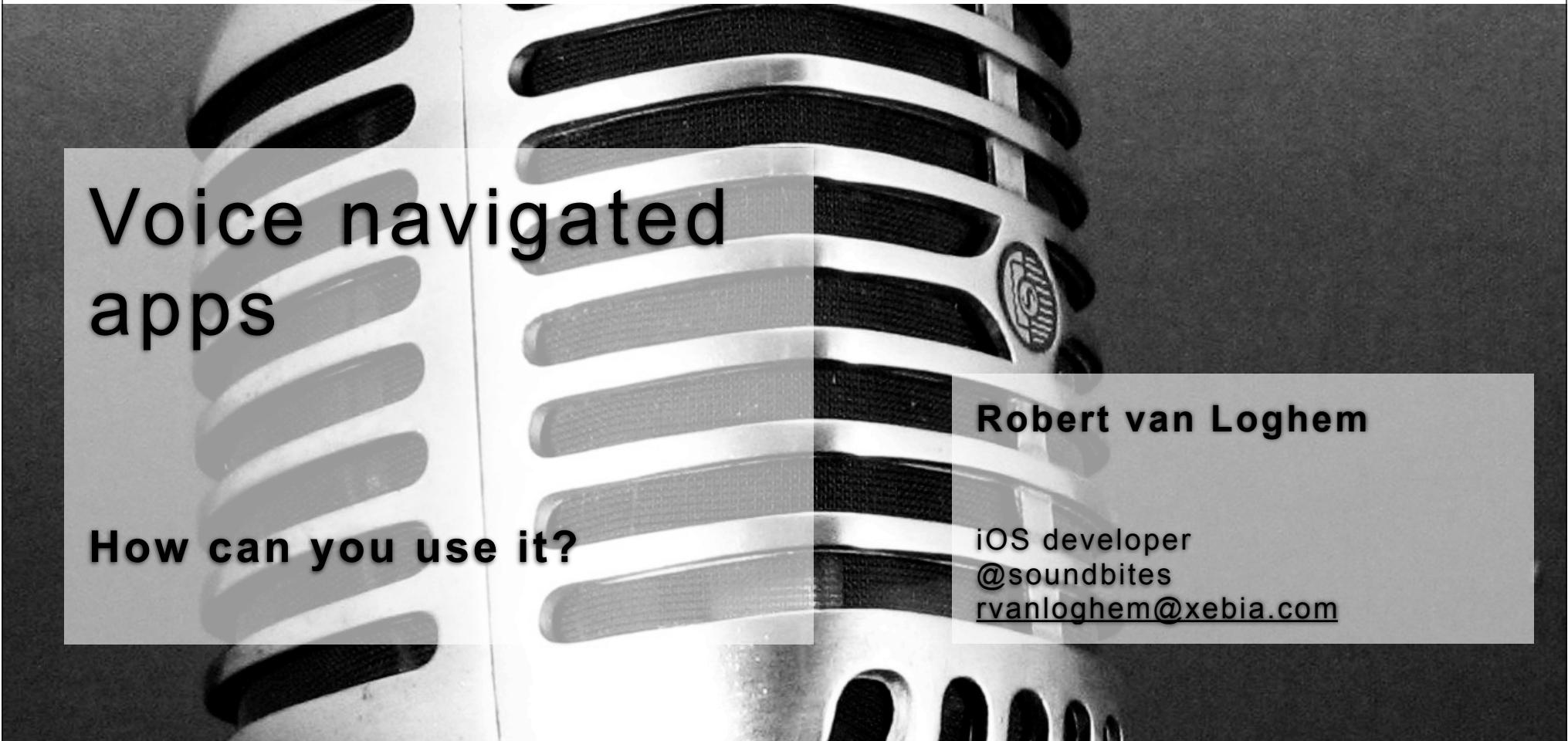
and they are happy!



Think about  
how you can  
use voice and  
try it out...

You can get **fast** results in very little time





Voice navigated  
apps

How can you use it?

**Robert van Loghem**

iOS developer  
@soundbites  
[rvanloghem@xebia.com](mailto:rvanloghem@xebia.com)



What does it take  
to make your  
app  
**Voice controlled**



Determine how  
you want your  
users to **interact**



Hint go for;  
**hard common tasks**





**Determine vocabulary**  
and make sure it  
resembles words  
used in your app





Make sure  
your users get  
**help**

What can they say...

Apple's Siri is a bad  
example...



When using TTS  
determine  
**tone of voice**  
when responding

A bank should be sincere,  
a youthfull company  
should use a youthfull hip  
voice



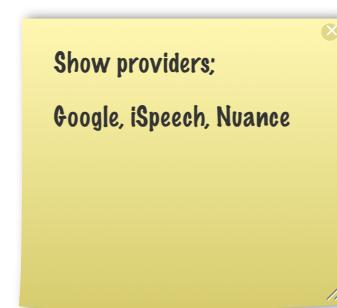
Most go for **queries**  
some for **dictation**

There is a difference in  
solution

Queries can be handled  
on-device, dictation you  
need more processing and  
is handled off device



Choose a  
**solution** to  
convert *speech*  
*to text* and vice-  
versa





When using  
queries, use  
**key-word natural  
language**  
recognition





Don't do  
**conversations**  
yet!.....wait! until the  
industry provides  
frameworks

