Designing a voice application
Top tips from UK voice producers

Nicky Birch | Lead Producer
Henry Cooke | Creative Technologist

2 May 2019
For the past three years, BBC Research & Development has been exploring conversation design and spoken interfaces. We’ve made a number of prototypes, published design guidelines and user test studies and published skills. We’ve also created our own conversational media authoring and publishing tools, collectively called Orator, which have gone on to support the BBC Kids Skill in production.

In February 2019, we invited a number of British voice producers to a workshop to discuss the challenges of creating voice applications for existing smart speaker platforms – currently Amazon Alexa and Google Assistant. The producers came from major UK publishers like The Guardian and Penguin Random House, start-up voice agencies like Opearlo and Labworks as well as some independent conversation designers and the National Health Service.

Creating high-quality voice applications and interactive media for smart speakers contains a tough set of challenges. We wanted to bring a panel of experts together to outline some of these fundamental challenges in designing for voice and share their top tips on how to tackle them.

If you are beginning to create a voice skill or application, especially one that uses audio and video rich media, you might find these tips useful.

This session focused on the practicalities of designing and producing an application – we decided to leave the equally knotty problems of monetisation, retention and discoverability for another day!
Starting out

Before even beginning to sketch out user flows and scripts, it’s important to spend some time thinking about your application’s core experience.

- **Do you have a clear set of goals** – business, UX, or both – to design towards? What’s in it for the user?

- What is the **core mechanic of your experience**? This could be a game mechanic or a conversational moment. Working this out and testing it early helps you know if you’ve got a compelling experience with repeat value.

- **Examine all the possible types of interactions** available on your chosen platform. What can the user do? How can you design around those affordances? Speech is just one channel of interaction – images on a nearby screen, interactions on a phone and many others can all be part of an experience.

- **Be careful when re-using existing web content**. Often, this loses a lot of context by being lifted from its original home. If you’re going to re-purpose content, think about how to rebuild that context.
Testing the idea

Trying an idea out as a working application can be expensive and complex to iterate. There are many ways to determine whether an idea will work well as a voice application before starting to write any code.

- **Storyboarding.** Draw a rough storyboard to outline what the user will experience while using your app. Using sticky notes as frames in your storyboard makes it easy to add or remove frames as you show your storyboard around your team.

- **Role playing.** Write a sample script and read it through, with one person playing the part of the user and another playing the part of the smart speaker. Emulate the limitations of the device: no eye contact, and remember that humans can be much more forgiving in conversation than an assistant. Be as dumb as a device!

- **Build a Wizard of Oz prototype** and test whether audiences respond to your flow, script and mechanic.

- **Build a working prototype.** If your skill is relatively simple, a quickly-built version with a synthesised voice could be enough to test the core of your idea. This can be great for demoing your idea to stakeholders.

- **Test individual interactions.** If the core of your experience is a repeatable interaction, like a quiz, then this is the part that can be tested in isolation. For example: do yes / no interactions work, or does the user need to give more specific answers?

- **Think about the user’s surroundings when using your skill.** Will they be in the kitchen, an office, in a car? Will they be with other people, or alone? Life is full of distractions and it’s worth bearing this in mind when testing. Don’t assume that you’ll always have someone’s complete attention, and test for this.
Scripting and flow design

As a conversation designer, your job is to create a natural interaction between computer and human. Even slight errors can make the user feel frustrated and less likely to proceed in your application.

- **It’s hard to separate script and UX design.** The role of a conversation designer is part scriptwriter and part user experience designer. Radio producers and game designers can also offer expertise in writing for audio and interaction.

- Many voice designers use off-the-shelf flowcharting software to design their basic flow. **Don’t fall into the trap of making the flowchart perfect!** It’s a tool to help you think, don’t get caught up in the detail.

- **Avoid overloading the start of the skill with instructions** – as a general rule, people can only remember two things at any one time. Spread instructions through an experience, and save them until the moment they’re needed.

- **Error handling** is a good time to introduce more instructions.

- **Prime the user with what is coming up**, so they know what to expect, when to pay attention and when to interact.
• **Create clear calls to action when the user is required to speak.** For example: a particular sound cue or a different voice can indicate when user input is expected.

• **When designing the vocabulary** you expect people to use when talking to your application, think about how words sound when spoken. Using words that sound very different to one another reduces the chances of utterances being mistaken for one another by the system. This is especially true for multiple-choice selections.

• **Visualising the conversation flow for designers and senior stakeholders is key.** This is where the flowchart can come in handy but even this can sometimes be unwieldy, some designers brought their ideas to life by recording role playing demos of the interactions and providing visual accompaniment in the style of a text message conversation.
Some tips for scripting

• **End sentences with a clear question.** For example: “are you ready?”, not “tell me when you’re ready.”

• When offering a choice, **don’t use more than three options.**

• **Don’t keep people stuck in error loops** - it’s better to progress on to something you didn’t ask for than to get stuck: “I didn’t catch that but let’s have another question.”

• **Don’t ask rhetorical questions.**

• Don’t tell people to “say this” or “say that” to do an action, **just ask the question.** For example: don’t tell people to “say continue to go on”, just ask “Would you like to go on?”

• **To stop Alexa triggering herself** if you use the word Alexa in the script, you can use the following techniques:
  • Remove 3500 to 6000 Hz
  • Speed up the word Alexa
  • Shorten the time between Alexa and the word before it, so they blur together
Creating a complex story

Designing a long-form interactive story with many branches can get very complicated very quickly. Branches may carry a number of properties, including images and buttons as well as audio assets, script and text.

Keeping track of user state often becomes necessary, storing which parts of a story a user has already heard and where they are in the flow. Reasoning about all of this complexity and gaining a ‘birds-eye view’ of the story becomes a significant challenge for any conversation designer.

- If your flow becomes **too complex to think about easily**, it’s possible that it’s **too complex for purpose**. Users may lose interest in a story that is laborious to discover, and even those that make it all the way through will only see a percentage of a complicated branching story.

- **Think about how to signpost content** in a complex story when most users can only remember two things at once.

- Over a certain size, **some parts of a story will always be missed**. Paths can be tweaked - and edits made - by using analytics once the story is live.

- **Video game design can teach us a lot here**. We can inform users of their progress, how much of the story they have ‘completed’ or how they have performed compared to others.
Using professionally recorded audio can make a voice application feel polished and high-quality, but simple changes to the flow or script can have a significant effect on your workload.

- **Leave studio recording of actors to the very end of your production process.** This gives you plenty of time to make adjustments to the script without incurring unnecessary studio costs.

- **Use a synthetic voice while honing the script.** You can later move to placeholder human voices – most often colleagues recorded locally – to avoid studio overhead when recording corrections.

- **Factor in re-records and pick-ups to your studio time** and project timelines.

- **Using recorded voices can add delays to a project** – recording, testing, re-recording and mastering all take time. Pay careful attention to how these parts of a project are managed to keep things on track.
Asset Management

If you’re building an application that uses a lot of media files – audio, images, video – things can become unwieldy, especially if you need to revise those assets when a script is changed or interactions are tweaked. There are a number of commercial voice publishing tools emerging, but at the time of our workshop there was no single off-the-shelf solution for keeping assets synchronised between a script, files on disk or in cloud storage, and application data.

• The most basic way to tackle this is by using a sprawling master spreadsheet that keeps track of file versions and script references. None of our producers recommended this approach, but some admitted to using it.

• A cloud based database / spreadsheet tool like Airtable is a minor improvement to this approach.

• Smaller projects, or those run by a single developer, have used JSON files to keep track of asset data.

• Larger companies generally used a bespoke content management system (CMS) which allowed non-developers to design, build and test applications. Some companies had designed their own CMS to fit their own needs (Including the BBC’s Orator) but off-the-shelf systems have also emerged. We expect this tooling to improve and become more widely available in the coming year.
The group agreed that the perfect publishing tool would include some analytics to enable tracking the effectiveness of content changes while a product is live.

For companies which publish existing content on other platforms (e.g. broadcasters) there are a few specific points:

1. Your asset management system needs to ensure a clear distinction between 'pre-existing audio' and 'bespoke voice audio' as mix-ups could be confusing for audiences.

2. When transferring existing web or app content into voice assistants, it can lose much of its context. There are online tools available that can summarise longform text content, but this still needs manual oversight to ensure it sounds conversational.
Mixing and Mastering

All of our experts have experienced problems in audio levels when mixing and mastering audio for smart speakers.

- **Loudness levels in your own audio can be hard to balance against system voices.** The best test is your ears – try different levels of normalisation against system prompts until you find a good level for your material.

- **When producing for Google use OGG files.** This will get you much better quality at smaller file sizes – which are essential for fast responses to the user.

- **Mix for the smart speakers themselves.** The best way to guarantee good sound quality is by listening frequently on your target device.

- When mixing audio, **use noise gates to silence sounds below a given volume floor.** This helps to filter out background noise in your recordings.

- **Filter out upper and lower audio bands** for rumble to give clearer sound.

The Guardian Voice Labs have written some more **useful tips** about mastering for smart assistants.
Thanks...

This document is a collaborative effort between several teams working on voice applications. We would like to thank the following for their time, advice and top tips -

*The Guardian Voice Labs*
*Opearlo*
*Labworks*
*The NHS*
*Penguin Random House*
*Hugo.fm*
Michelle Feuerlicht
Christian Menniss
Oscar Schafer
Martin Johnson