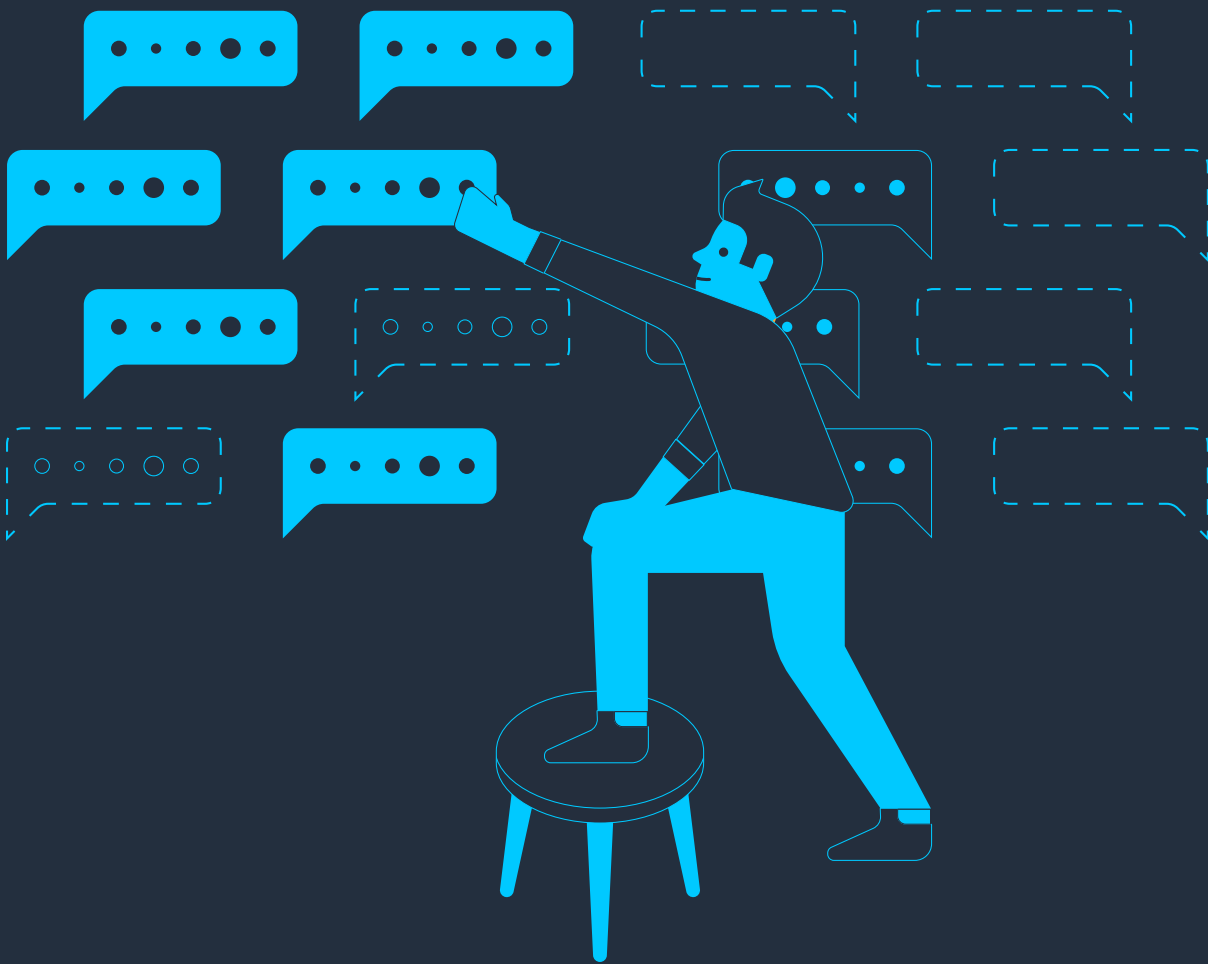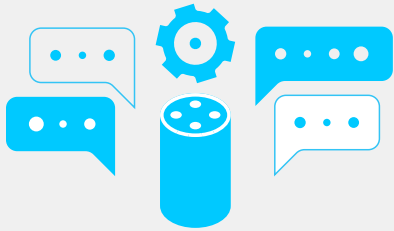# Create, Manage & Update Content for Voice

Introduction

Content is at the core of digital products, including Alexa skills. **A content management system (CMS) is an application that is used to create, manage and update such content.** CMSs are widely used for web, mobile, and enterprise content management and are equally applicable to managing voice experiences and Alexa skills.

The top reasons why brands implement a CMPS for their skills are:

1. Collaborate on content
2. Enable personalization
3. Empower non-developers to make instant updates
4. Maintain consistency or introduce variety
5. Simplify localization

In this Content Management Spotlight, you'll learn about:

⊘ Why is VUI **different?**

⊘ What content can be **easily changed?**

⊘ **Best practices** for skill content management

⊘ **Common pitfalls** in skill content management

⊘ Skill content management **tools & resources**

Why Is VUI **different?**

Voice UI (VUI) is fundamentally different from Web and Mobile UI (or GUI). As a result, the tools required for managing, analyzing, optimizing and personalizing the voice experience require a fundamentally different approach from the ones that are commonly used for Web and Mobile.

When glancing at a screen, users scan across a wide range of visual information (layout, text and images) and selectively pick out what they want to focus their attention on. In contrast, when interacting with a skill, users listen to a single stream of voice content, which means that getting that content right is absolutely critical for keeping users engaged. Therefore, it is essential to apply best practices when creating and managing voice content.

There are five fundamental differences between VUI and GUI:
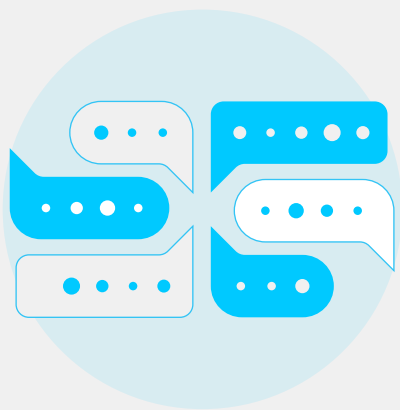
1. VARIETY VS. CONSISTENCY

Web and mobile UI are built on consistency — users scan a screen and expect to see consistency in how the content is laid out. In contrast, consistency in a voice UI leads to a boring, unengaging and repetitive user experience.

One of the defining characteristics of human conversation is the unpredictability of the dialogue. A conversation can take many twists and turns and go down multiple paths to get to the same endpoint. Keeping the dialog fresh by delivering variations in the responses allows the skill to sound less robotic and predictable. Variety is what keeps users engaged and coming back for more.

2. PERSONALIZED FEED VS. GENERAL SCAN

Web and Mobile UI require users to do more work sorting out the relevant content on the screen or in an app.. Skills, on the other hand, deliver a single audio stream of content to the user that needs to be contextual and personally relevant, or they'll tune out.

Multi-modal UIs (e.g. Echo Show) — those that combine voice with visual content — add further challenges and complexity. Fortunately, there are optimization and personalization solutions available that determine the right level and structure of information to present on screen, taking into account both what's being spoken as well as the capabilities of the device.
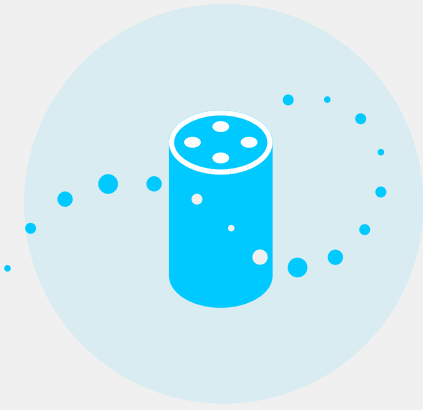
3. FREE-FORM VS. PRE-EXISTING STRUCTURE

The appeal of VUI is that users can be in the driver's seat of the experience. Through their questions and prompts, they define the structure of each session. This is in stark contrast to Web and Mobile, where users navigate a pre-existing structure of links and touch controls.

The free-form nature of voice interactions means that skills need to be built in such a way that allows them to switch among multiple contexts based on how their users choose to take the conversation.

4. WHAT VS. HOW

While the exact same information can be conveyed on Web/Mobile and Voice, skills have the unique advantage (and added complexity) of defining how that information sounds when it is delivered. Through a combination of TTS with SSML and pre-recorded audio, skills can fine-tune the sound, delivery and even emotion of the content expressed. It means that authors have a higher level of control over the user experience, but it also means that they have to invest in getting it right.
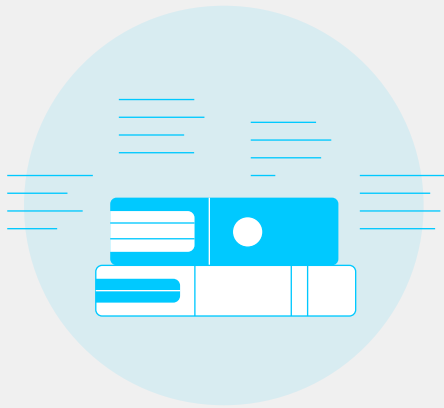
5. DIALOG VS. NARRATIVE

The content that exists in a skill is in a dialogue form. Ideally, the content comprising the responses of the skill should be largely in the form of small snippets that are combined and delivered based on context.

On the other hand, Web and Mobile have large blocks of content that provide structure to the product and in general are unchanged across time or users.

4

What Content Can Be
Easily Changed?

Unlike updates to the interaction model (utterances and intents), which require skill recertification, skill responses can be easily updated. The only restriction to doing so is the tools and methods used to make changes.

Responses are the information the skill shares with users. It is therefore very powerful to be able to keep that information up-to-date, in order to keep users coming back for ongoing interactions.

Let's say that the skill publisher has new content it wants to immediately share with customers. For example, it could be newly recorded audio, a marketing campaign, or an updated product offering. Getting that content out to consumers as quickly as possible is crucial to the skill's relevance, engagement, and retention.

For more information on what updates can be made to a live skill, please read the **Update Live Skills in Minutes** post on the developer blog.

An added challenge is that more often than not, the content owners are not developers and therefore need tools in order to push fresh content without engaging the engineering team and without risk of breaking the skill. This is where a Content Management System (CMS) comes in handy. Let's take a look at how to get started.

Best Practices for
Skill Content Management

1 EXTERNALIZE CONTENT STRINGS FROM CODE

The first step towards effective voice content management is to separate content from the application's source code into language-specific resource files. Having separated content opens up significant possibilities to updating it without having to dig into the source code and change the same item in multiple locations. Open-source libraries that offer a good starting point are available here:
Jargon SDK for Amazon Alexa

2 MAKE CONTENT VARIED AND DYNAMIC

There are two separate concepts that make skills more engaging for end users.

The first concept is the use of variants. Without variants, skill responses can become predictable and frustrating to hear repeated over and over again. Each skill response can have multiple variants — different ways of providing the same information to consumers. The more variants a skill has, the more conversational and richer the user experience will be. Learn more here.

The second concept is to make content dynamic. Dynamic content enables two specific scenarios — the ability to adjust the content of the response based on context (e.g., who the user is, how they've previously interacted with the skill, etc.) and the ability to push content changes without requiring a redeployment of the skill.

Commercial tools like Jargon make it easy to introduce variants and enable dynamic content.

3 ENABLE NON-ENGINEERS TO UPDATE CONTENT

Implementing a CMS with a Web interface allows non-technical content owners to update the skill without the risk of inadvertently breaking it. Removing the dependency on the development team for content changes dramatically unlocks the publisher's ability to push out updates and keep the skill fresh.

Available commercial solutions also support content versioning and collaboration features, making it easy for multiple stakeholders to work together on new content and ensuring that the right approvals have taken place before it is published to consumers. Also, in the event of a mistake, the right tools make it easy to revert to a previous version.

4

With just a bit of foresight and consideration on future **localization needs**, a skill can be built in a way that dramatically eases the expansion to new locales. Read about some of Alexa's best practices on localization **here** and **here**. In addition to building on Step 1 above, developers should use the **ICU format for parameters**, keep track of currencies and other locale-specific metrics, and take note of local or cultural references that don't easily apply across geographies.

5

UTILIZE MULTIMODAL ASSETS

The power of Alexa skills is that in addition to voice, customers can use the Alexa Presentation Language (APL) to augment the user experience with rich visual content. Video, audio, images and card content are all an essential part of the skill and should be managed as such. Consider having a single place where non-text assets are managed and referenced, in order to facilitate updates. For example, if a new video is available to replace an existing asset, it's ideal to have a single location in the CMS to swap that asset, reflecting the change across the skill, whenever that video is referenced. Commercial solutions already exist making such asset management dramatically simpler.

6

ANALYZE, LEARN, OPTIMIZE

As with any other product, a skill should be an ever-evolving product that gets analyzed and fine-tuned over time, reflecting the publisher and its customers' evolving needs and preferences. With voice interactions in particular, there are endless variations of utterances and intents that customers might try, so it is critical to learn from usage data and update the skill accordingly.

A number of **Analytics tools** can provide the data and insights needed. In addition to the analytics around user input and usage, it is important to track the performance of different content versions. For example, specific variants of a skill's response might resonate best with users of certain demographics or characteristics, whereas other variants perform best with other subsets of users. Analyzing the performance of content across user cohorts can help fine-tune the experience at the user level. Such content-related analytics are available in some of the CMS tools currently on the market.

Common Pitfalls in
Skill Content Management

1

RELYING ON LIMITING TEMPLATES

Templatized solutions for skills are a fantastic way to get a quick start. They can be used in production for simple skills or basic prototypes to test an idea. However, for skills of any meaningful complexity, templates can be limiting.

Given the multitude of products and tools available, it is important to find that right match that empowers the team and speeds up skill development/management, without limiting the user experience. A good litmus test is to validate that the development team can still have full control over the logic of the skill and is not limited by a set of templatized interactions, while non-developers can freely update content.

2

WRITE AND FORGET

When launching a new skill, it is critical to plan for the ongoing content refresh required to keep the skill relevant. Content-driven products are only as useful as the quality, relevance and freshness of their content.

3

SKILL LOCALIZATION IS NOT A ONE-TIME EFFORT

Publishers have an opportunity to reach global audiences with the ever-increasing geographic expansion of Alexa. Teams often start by writing a skill in one language and then decide to localize to additional ones.

There are several complexities that come with skill localization:

a. There needs to be a mechanism in place to extract all of the strings (utterances, responses) that need to be localized.

b. Translators need to have enough context to know how best to localize the content.

c. Translation is not always a 1:1 mapping across languages. For example, an utterance in English can be mapped to several equivalent utterances in French. With the right tools in place, localization offers a great opportunity to capture this variety, as it will improve the usability of the French skill.

d. Machine translation alone does not result in a suitable, fine-tuned localized version.

e. Localization is not a one-time effort. As the original version of the skill changes, there needs to be a process in place to keep the localized versions updated.

There are localization tools that are purpose-built for skills and can help streamline the manual parts of localization, as well as enable seamless localization refreshes as the original skill changes. Jargon provides good options for localization tools that translators can get direct access to, or the option to use one of their partner localization providers.

Tools and **Resources**

Constructing and managing effective voice content requires a skill to balance among numerous constraints and priorities, such as the length of the response, fulfilling the user's immediate need, and setting the stage for future interactions. Satisfying these competing demands requires systems that are purpose-built to account for the needs and nuances of voice; simply repurposing existing web or mobile content management and optimization techniques is inadequate.

Here are a few suggestions:

Jargon: CMS for Voice

Jargon SDK: Open Source Library for managing content

Alexa Blog: How to tailor your voice experience for global audiences

Alexa Blog: How to localize your Alexa skills