<u></u> () () () () () () () () () ()	<u></u>	00000000000000000000000000000000000000	000000000000000000000000000000000000000	୦୦୦୦୦୦୦୦୦୦୦୦୦୦ ଜଣ୍ଣଜ୍ଞାନ୍ତ୍ରାର୍ଭ୍ବାର୍କ୍ତ୍ର	00000000000000000000000000000000000000	000 00000000 00000000	00000000000000000000000000000000000000		000000000000000000000000000000000000000	000000	දා දා දා දා දා දා	ଐ ଓ ଅନ୍ତି ଅନ୍ତ ଅନ୍ତି ଅନ୍ତି ଅନ୍ତ ଅନ୍ତି ଅନ୍ତି ଅନ୍ତ	ම අ අ අ අ අ අ අ අ අ අ අ අ . ම අ අ අ අ අ අ අ අ අ අ අ අ අ	00000000000000000000000000000000000000	\$. 산 산 산 산 산 산 산 산 산 산 산 산 산 산 산 산 산 산 산 수 수 수 수	භු භූ භූ භූ භූ භූ භූ භූ භූ භ භ භ භ භ භ භ භ	~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~	ᅉᅉᄻᅉᄻᅉᄻᅉᄻᅉᄻᅉᄻᄻᄻᄻᄻ ᅓᅓᅒᅒᅭᇑᇑᇑᇑᇑᇑ	ස් ස් ස් ස් ස් ස් ස් ස් ස් ස් ආ වා	\$\$\$\$\$\$\$\$\$\$ \$\$\$\$\$\$\$\$ \$\$\$\$\$ \$\$\$ \$\$\$ \$\$\$	ゆ ゆ ゆ ゆ ゆ ゆ ゆ ゆ	. ආ ආ ආ ආ ආ ආ ආ ආ ආ ආ ආ ආ ආ ආ ආ ආ	(1) (1) (1) (1) (1) (1) (1) (1) (1) (1)	
			Ô	୍କ କୁ ଅକ୍ଷ ଅ	\bigcirc			わうこう)) (i)		() () () () () () () () () () () () () (27 27 27	් දුරු දුරු ද	යා දා දා දා දා දා දා දා දා දා	ර ි ි ි ි ි ි ි ි ි ි ි ි ි ි ි ි ි ි ි	ŝ					

WHAT CONSUMERS WANT **IN VOICE APP DESIGN**

AN EMPIRICAL STUDY NOVEMBER 2019

A RESEARCH COLLABORATION BY:



voicebot.ai Voices.com pulselabs

TABLE OF CONTENTS

Introduction Page 3

Voice Provenance & Gender Preferences Page 4

User Preferences & Recall with Variable Lengths Page 13 Conclusion

Page 18

ABOUT VOICEBOT

Voicebot produces the leading independent research, online publication, newsletter and podcast focused on the voice and Al industries. Thousands of entrepreneurs, developers, investors, analysts and other industry leaders look to Voicebot each week for the latest news, data, analysis and insights defining the trajectory of the next great computing platform. At Voicebot, we give voice to a revolution.

METHODOLOGY

240 panelists were assembled between July 2 and July 8, 2019 in order to gather opinions on feedback for a series of voices, both human and machine generated. These panelists were separated into 3 groups, which all listened to the same first four voices before listening to a unique voice for each group. Additional segmentation was done for the speech duration and call-to-action recall testing.

ABOUT VOICES.COM

Voices.com is the largest global marketplace for audio and voice over services. With clients in 160 countries, and more than 200,000 voice actors speaking over 100 languages and dialects, Voices.com helps businesses solve their voice needs, at today's speed of production.

ABOUT PULSE LABS

Pulse Labs is the premiere measurement, analytics, and testing platform for voice. It's products and solutions help marketing and product teams better understand and enable deeper consumer engagement with their brands and applications on voice platforms.

A SPECIAL THANKS

The research team would also like to express a special thanks to Skilled Creative and Meredith Corporation, publisher of Entertainment Weekly, for their active support of the study.



Voicebot, Voices.com, and Pulse Labs collaborated on a recent study that sought to put some hard data behind a series of voice user experience (UX) questions. There are plenty of opinions about voice UX best practices, dos and don'ts, pitfalls and opportunities. However, these statements that are so often made with such certitude that they are positioned as fact, are invariably subjective. While we cannot transform UX into a totally objective science, our hypothesis was that we could develop a series of tests to measure actual consumer preferences related to a series of questions.

- Do voice assistant users prefer human voices over synthetic voices and if so, by how much?
- Do users prefer male or female voices? Is that preference the same when considering human or synthetic voices?
- How does age of the user impact these preferences?
- How long is too long when delivering content through a voice assistant?
- Is there a difference in tolerance for the length of voice assistant content delivered by a human compared to a synthetic voice?

Our research considered these and other questions by presenting a series of exercises to a panel of 240 consumers managed by Pulse Labs. Some results will surprise no one while others introduce nuance into our understanding of consumer preferences when it comes to interacting with voice assistants.



There is a Strong Preference for Human Voices

It will surprise no one that our user panel expressed a preference for human voices over synthetic voices. Observers have long suspected that users preferred to hear humans. In our testing, human voices received an overall rating of 3.86 on a scale of 1.00-5.00 compared to 2.25 for synthetic voices generated by artificial intelligence. That difference reflects a 71.6% higher rating for human voices over the synthetic alternative.

Voice Preference Rating



VOICEBOT.AI

These differences are placed in starker contrast when you consider the relative responses of the evaluators in terms of favorable, neutral, and unfavorable ratings for the human and synthetic voices. Nearly 71% of study participants rated the human voices favorably compared to 12.5% unfavorably. Data from the synthetic voices is reversed. Just over 60% rated the synthetic voices unfavorably compared to 12.3% favorably.



Age Matters More Than Sex in Preferences

The preference holds across the three age groups evaluated which included 18-29, 30-59, and 60+. The 60+ age cohort had the largest disparity between human and synthetic voice options expressing a 91.4% preference for the former. They were followed by the 18-29 group with an 80.1% preference for human voices. The 30-59 cohort was the most tolerant of synthetic voices but still expressed a 66% preference for human voices over synthetic.



Voice Provenance Preference by Age of Evaluator



When it comes to the sex of the evaluator, there was almost no variance in the results. Women preferred human voices 72.9% more than synthetic while men expressed that same preference at a rate of 69.6%.

Voice Provenance Preference by Sex of Evaluator



What Consumers Said About Synthetic Voices

The initial part of our study set out to evaluate the relative preference for human or synthetic voices and compare those expressed preferencs based on age and sex of the evaluator. However, we also wanted to know whether the gender of the voice assistant impacted the sentiment and, if so, by how much.

VOICE	CATEGORY
1	Synthetic Female
2	Human Male
3	Human Female
4	Synthetic Male

In the table we show that voices 1 and 4 were synthetic female and male respectively, while voices 2 and 3 were human male and female respectively. The dominant adjectives used to describe voices 1 and 4 were robotic, monotonous, monotone, and boring. Other common descriptions are included in the word clouds below.

VOICE 1 - Synthetic Female





Some commentary about Voice 1 included that it, "sounded mechanical," "the cadence was hard to follow," "it was boring," "sounded flat and uninteresting," and "too monotonous." Voice 4 had some similarly negative commentary such as "I didn't like how cold the voice sounded and distant as well," "insincere-sounding, a very strange tone," "it was whiny and put me on edge."

These are clearly not the type of associations that a voice app developer wants attached to their persona. However, there were a minority of users that expressed favorable thoughts on voices 1 and 4 making comments such as, "I liked that I was able to clearly understand the message the voice was trying to convey," and "it was the easiest to listen to." There were far more negative comments but it is worth noting that some consumers may in fact prefer characteristics of synthetic voices.

What Consumers Said About Human Voices

By contrast, the human voices had far more positive adjectives associated with them. For example, consumers in the study said about Voice 2:

"I liked how pleasant the voice was. It felt like a friend talking to me."

"It sounded like a radio personality. It was smooth and sexy."

"There was excitement in his voice which got me excited about the content being shared."

The most common words for Voice 2 were enthusiastic, upbeat, radio, energetic, and exciting. Voice 3 had a similarly positive reception among the study participants.

VOICE 2 - Human Male



VOICE 3 - Human Female



It was well received overall with consumers saying Voice 3 was "natural, happy, bubbly," "sounded like a friend telling me about the new album," "it was cool and alive in a soft voice." The most common adjectives included "friendly," "youthful," and "natural."

There were a few users that had negative reactions and offered comments such as "it was whiny and put me on edge," "she sounds like a child. If your target is 11 year-olds, perhaps she's the one to go with."

There is a Small But Measurable Preference for Female Voices

We can see from the data that human voices are strongly preferred overall and by demographic breakdowns for age and sex. When it comes to the gender of voice assistants there are also preferences expressed by our evaluators but the differences were not as stark.

This has been a hotly debated topic for a number of reasons. One assertion that is often challenged rhetorically is Amazon's claim that it chose to make Alexa's voice female based on preferences expressed by early testers of the assistant. Our study discovered a similar preference for female voices when they are synthetic. But there is a wrinkle in this story.

The first preference that we can see is the disparity between male human and synthetic voices compared to female human and synthetic voices. Male human voices were preferred by evaluators over synthetic by a margin of 84.7% while female human voices were preferred at a rate of 60.5%.



This result means the evaluators were more tolerant of female synthetic voices relative to the human option compared to male voices. The reason for this is unclear. It may be that the widespread use of female synthetic voices in the U.S. with Siri as a default female persona and Alexa as only a female persona, has conditioned users to be more accustomed to assistants that sound like women. It could also be that the most famous of science fiction voice assistants, HAL 9000 from the movie 2001: A Space Odyssey, has conditioned users to be wary of voice assistants that sound like men. Whatever the reason, in our test group, the disparity was material. At least it was material for synthetic voices.

The evaluators expressed a 12.5% preference for synthetic female voices over synthetic male voices. However, there was just a 2.3% preference for male over female when the voice was human. This is an intriguing finding. The preference of male over female human voices for the full panel narrowly favored the masculine.

Gender Preference for Voice Assistant Personas



There will always be a degree of subjectivity when assessing human preferences for things that are not all exactly alike. There are no two human or synthetic voices that are truly identical so subjective preferences can creep into the data. It is logical to argue that for human voices, that male and female are so close that for a sample of 240 it is within the margin of error. However, you would also acknowledge using this same logic the preference for female voices when both are synthetic is well outside the margin of error. The results seem to confirm Amazon's stated analysis. However, we wanted to look into this finding in more detail to see if there was a difference in preferences expressed by men and women. The data say that both men and women prefer female synthetic voices. Women show a preference of 11.9% while men come in at 14.3%.



Synthetic Voice Gender Preference By Sex of Evaluator

The female gender preference for synthetic voices also shows up in each of the three age groups in our study. Participants that were over 60 years old showed the strongest preference at over 23%. They were followed by the 18-29 year-olds at 13.4% and 30-59 year-olds at 11.8%.



Synthetic Voice Gender Preference By Age of Evaluator

When we look at gender preferences for human voices the difference narrows to almost even but tips slightly toward males. However, that overall figure is driven by women evaluators as men expressed a slight preference for a female human voice. Women chose the human male voice at a rate 3.8% higher than a female voice whereas men favored female voices by 0.3% which is close enough to consider it no preference.

VOICEBOT.AI



Human Voice Gender Preference by Sex of Evaluator

Findings for Voice Provenance & **Gender Preferences**

In terms of voice provenance, the preference for human over synthetic voices is conclusive. There are some evaluators that expressed a preference for synthetic voices so it is not as if the conclusion is unanimous. However, voice app publishers should know that use of human voices is likely to be received more positively by users.

When considering the gender of the voice, the data suggests consumers will respond more positively to female voices if using a synthetic voice and either gender is likely to be received about equally if a human voice is selected. Of course, there may be other factors that influence the gender preference for a specific voice app. The type of voice experience, the nature of the content, and context in which it is used could all influence user preferences. This study did not attempt to consider those other variables, but voice app publishers should keep them in mind when designing their user experience.

\$\$\$\$\$ ග හ හ හ ଭୁଷ୍ଠଭୁଷ୍ଠଭୁଷ୍ଠ ଭୁଷ୍ଠଭୁଷ୍ଠଭୁଷ୍ଠ ෯෯෯෯෯෯෯෯෯෯෯෯෯෯෯෯෯ ෯ ෯෯෯෯෯෯෯෯෯෯෯෯෯෯෯෯෯෯෯෯෯෯෯෯෯ ඟි ගි ගි ගි ගි \$\$\$\$\$\$\$\$\$\$\$\$\$\$\$\$ \$ \$\$\$\$\$\$\$\$\$\$\$\$\$\$\$\$ \$\$\$\$\$\$\$\$\$\$\$\$\$\$\$\$\$ **USER PREFERENCES & RECALL WITH VARIABLE LENGTHS**

There is a Strong Preference for Shorter Dialogue

The next item we assessed was how users responded to the length of content presented in typical voice assistant scenarios. There are occasional claims in the UX community about what makes good conversation and how long voice assistant dialogue can be before it becomes tiresome to users. However, these claims seem to invariably be unsubstantiated and based solely on opinion or anecdote.

For this testing, we first employed the Entertainment Weekly Alexa skill implemented in four variations based on content length. These variants were tested among four different groups. The variations were comprised of an introduction (long or short) followed by a list (long or short).

GROUP	INTRO	LIST				
One	Long	Long				
Two	Short	Long				
Three	Long	Short				
Four	Short	Short				



User Content Rating Based on Dialogue Duration

The first point that jumps out is that Group Four, with the short/short configuration has about 18% of respondents saying the content presented was "too short" and no one in the other groups assigned the "too short" label at all. Another finding you might notice is that the ratio of "neither too short nor too long" is almost identical across Groups Two, Three, and Four, all falling in a range of 50-56%. Groups Three and Four both had a short list configuration and they fell in the much tighter range of 54.5-55.6%.

Only 18.2% said that the long/long configuration was "neither too short nor too long." The remaining 81.8% concluded the dialogue was "too long." This was three times higher than the Group Four short/short configuration for the "too long" designation. Groups Two and Three were 1.63 and 1.83 times higher for "too long" ratings than Group Four.

The data also show that the long introduction was more tolerable than the long list. There is general agreement that lists become tedious in spoken dialogue. We see in the two balanced Groups, Two and Three, where we have a long and short combination, that the short list is 11% more favored as "neither too long nor too short." We also see that the "too long" designation nearly doubles when going from a long intro with a short list to a long intro with a long list. Combined length surely is a key factor, but the longer list is the factor with more user weighting.

Users Rate Skills with Shorter Dialogue Higher

Another point of note for developers and publishers is the potential impact of dialogue length on voice app rating. The Group Four short/short received a 16% higher rating than Group One long/long with 3.80 to 3.27 respectively. Groups Two and Three which had the balanced long and short combinations had ratings within 1.5% of each other though they still trailed Group Four by a noticeable margin.





Short Dialogue Preferences and Recall Extend to Voice Provenance

Our second set of tests reverted back to evaluating the impact of human versus synthetic voices. A short dialogue from a synthetic Poly voice of 25 seconds outperformed the same voice with a longer , 49-second, duration. The shorter variant of Voice 5 was ranked "about right" by 66.4% of users compared to 42.9% for the longer Voice 6. There was an even greater disparity between Voice 5 and 6 when it came to the combined "too long" and "far too long" designations which totaled 29.9% and 55.3% respectively. We see once again a confirmation that shorter dialogue leads to better user ratings of the experience.



User Preference by Length of Voice Assistant Dialogue

We also see that there is more tolerance for human voices with longer dialogue than a similar duration synthetic voice. This result can be seen when comparing Voices 6 and 7. Voice 6's long synthetic dialogue received a combined "too long" and "far too long" designation of 55.4% compared to 37.5% for Voice 7.

Another interesting finding is that the Short Poly voice (Voice 5) and the Long Human voice (Voice 7) are nearly identical in user ratings. Only about five percentage points separate these two voices for the "about right" figures and the combined "too long" and "far too long" results fall within eight points. With that said, this is an instance where a synthetic voice actually beat a human voice. The key variable that drove this outcome was shorter dialogue.

User Call-to-Action Recall is Higher with a Human Voice

Finally, we evaluated the impact of human and synthetic voices on user recall of a call-to-action. This was conducted in conjunction with the previous test by adding a call-to-action in the dialogue. The human voice with a long dialogue was more than twice as successful in aiding user recall than either a short or long synthetic voice.



Information Recall by Voice Type and Dialogue Duration



In fact, duration didn't seem to have an impact on correct call-to-action recall. The long and short synthetic voice had about the same success rate. The key difference was in recollection that a call-to-action was included in the dialogue. Far more users recalled there was a call-to-action with the short dialogue compared to the longer version.

Interestingly, the user recall that there was a call-to-action present was nearly identical between Voice 5 with the short synthetic speech compared with the human voice (Voice 7) which had far longer dialogue. This outcome was similar to the dialogue duration ratings above.

We know of no previous study that has attempted to quantify consumer preferences about dialogue length and recall related to voice assistants. However, we welcome further studies to explore this topic in more detail. Voice user experience designers could benefit from more evidence-based discussions around these topics so they can better gauge the implications of dialogue copy and voice selection and match these variables to project goals.



The research shows that voice assistant users prefer shorter dialogue and human voices. User preferences were consistent across multiple tests. So, if voice app designers want to maximize their chance of user acceptance, then brief and human dialogue is a reasonable strategy. However, if for some reason a voice app publisher cannot have short dialogue then they could potentially offset negative user sentiment by switching to a human voice over. Or, if they have to stick with a synthetic voice, then keeping the dialogue brief is the best strategy for success.

Amazon and Google appear to understand at least some of these variables. Alexa and Google Assistant both bias toward shorter dialogues and chose to have a synthetic voice to provide maximum flexibility for rendering the widest variety of speech. Granted there are other variables at play and some of those may be equally important to consider as voice provenance, gender, and duration of the dialogue.

Voice user experience design has been around for decades, but it remains a young science when it comes to the latest voice assistant use cases. Whether you are launching new Alexa Skill, Google Action, Bixby Capsule, or you own assistant, we hope the data presented here offers you fresh insight into how design tradeoffs can influence your user experience.

In closing, thanks go to our co-sponsors Voices.com and Pulse Labs, as well as study supporters Skilled Creative and Meredith Corporation, publisher of Entertainment Weekly. Your contributions to the study design, data collection, and analysis were invaluable.

|--|--|







To request custom voice or AI industry research contact: info@voicebot.ai