Amazon Echo: A Benchmarking Model Review

Dietmar Jakob, Sebastian Wilhelm
Technology Campus Grafenau
Deggendorf Institute of Technology
Grafenau, Germany
<firstname>.<lastname>@th-deg.de

Abstract

Smart speakers are becoming increasingly popular. The market leader for smart speakers are the products of the *Echo* family from *Amazon*. There are currently 9 different models with different technical specifications available in Germany. With this paper, the models was benchmarked against each other in terms of (i) speech recognition reliability, (ii) output sound pressure and (iii) power consumption in a laboratory experiment. Previous works in this area has only considered individual models of the product family. Significant differences in speech recognition accuracy, output sound pressure and power consumption were identified between the models. In general it was observed, that the *Echo Show 8* model was the most efficient in terms of the above criterias.

Index Terms

Echo, Alexa, Voice User Interface, Speaker Recognition, Usability, Human Computer Interaction

I. Introduction

The global smart speaker market grew by 44.9% in the third quarter of 2019 compared to the third quarter of 2018. In the third quarter of 2019, 28.6 million units were sold globally. *Amazon* dominates the market alongside *Alibaba*, *Baidu*, *Google*, *Xiaomi* and others with 10.4 million smart speakers sold in the third quarter of 2019. Therefore, *Amazon* was able to increase its market share from 6.3% to 10.4% (Canalys, 2019).

The smart speaker family from *Amazon* is known as *Echo*-devices and relies on the voice user interface *Alexa*. In Germany there are currently (per February 2020) 9 different models available: (i) *Echo Show (2. Gen)* (ii) *Echo Show 8* (iii) *Echo Show 5* (iv) *Echo (2. Gen)* (v) *Echo (3. Gen)* (vi) *Echo Plus* (vii) *Echo Flex* (viii) *Echo Dot* (ix) *Echo Studio*.

The market leadership of *Amazon* and the variety of its avialable smart speakers, is the motivation to examine the individual products of the company in more detail from a technical point of view, as well as the reliability of speech recognition. The main contribution of this paper is, to examine the differences between the available models of the *Echo* family in the aspects of: (i) how precisely does it follow the recognition of the deposed voice commands (ii) , at what distance do the devices still understand the voice commands (iii) , what is the power consumption of the individual models (iv) , how does the sound pressure change as output volumes increase.

The remaining of the paper is structured as follows: The work begins with an overview of similar scientific studies in section II. The theoretical and experimental benchmarking of the *Amazon Echo* devices will be presented and discussed in section III before the work ends with a summary of the results and an outlook for future research in section IV.

II. Related Work

A. Speech and Voice Recognition

A key feature of smart speakers is the recognition of voice commands and the execution of corresponding actions. *Gao et al.* investigated the speech recognition and understanding of the *Echo* devices based on customer reviews at the *Amazon* website. They found that customers often complain about the precision of speech recognition and/or the ability of *Echo* devices to answer questions correctly (Gao et al., 2018). In contrast *Hellwig et al.* summed up that the recognition of the language was 'relatively good', however they although noted that *Echo* could not understand all commands so that the interaction is aborted (Hellwig et al., 2018). *Lopatovska et al.* found that for advanced users 89% of commands are understood, while for non-advanced users the rate is only 56% (Lopatovska et al., 2018). By asking 25 Japanese about their experiences with language assistants, *Yamada et al.* found that 50% had encountered problems with speech recognition (Yamada et al., 2018). Further investigation of *Shalini et al.* showed in an experimental way, that some words of 12 different sentences was misinterpreted. For example, commands containing the terms 'health', 'pulse', 'what', 'an hour', 'vacation', or 'early' were understood and interpreted as other terms (Shalini et al., 2019).

All the studies mentioned above came to the conclusion that the voice commands were not reliably recognized. However, none of them quantified the precision of speech recognition. Further, the individual *Echo* models were not examined and compared individually.

B. Power Consumption

None of the current Echo device own a power switch. Therefore, the device are continuously in a standby operating mode. Up to now, less investigations was made on analyzing the power consumption of individual Echo devices. *Lloyd* analyzed the power consumption data from the *Echo Dot*, *Echo (1st Gen.)*, *Echo (2nd Gen.)*, *Echo Plus* and *Echo Spot* models in standby-mode and while playing music. His findings are shown in Table I (Lloyd, 2018). His investigations lacks the models *Echo Show*, *Echo Show 8*, *Echo Show 5*, *Echo Flex* and *Echo Studio*. Furthermore, the methodology are not very resilient, since *Lloyd* only made a comparison in standby operation and music playback.

Action	Echo Dot	Echo (1. Gen.)	Echo (2. Gen.)	Echo Plus	Echo Spot
Standby	1.75	2.95	1.95	2.4	1.9-2.25
Playing music	2.1-2,4	3.1-3.4	2.4-3.4	3.0-4.3	2.6-3.2
Average Playing music	2.25	3.25	2.9	3.65	2.9

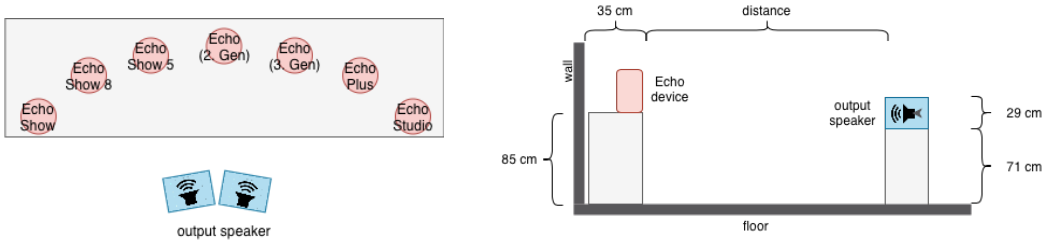
TABLE I: Power consumption data (in *W*) (Lloyd, 2018).

Frawley determined the power consumption of several *IoT* devices from different manufacturers. In the smart speaker category, he examined the *Echo Dot* and the *Echo Show*. Each smart speaker is prompted to read the weather, read the news, and play music every time it is used. Once a week, he asked the speakers to set a reminder for the future. Each speaker was also muted for a few minutes to see, if there are any noticeable changes in power consumption during the period where the device is not listening for. He noted that when the environment darkens, the LEDs of the devices light up brighter. This causes the *Echo Dot* to consume about 2200mW of power rather than its usual 1500mW (Frawley, 2018). The trial period of one week and the small number of different models do not allow meaningful conclusions to be drawn about power consumption under loads.

III. Technical Benchmarking of *Echo* Models

In order to benchmark the *Echo* models with each other in the fields of speech recognition, output volume and power consumption, all currently available models in Germany (per February 2020) was investigated.

In the following, first the general technical specifications of each model are presented in subsection III-A. With subsection III-B the power consumption of the devices is reviewed over a 24-hour period during which the devices received commands continuously. Next, in subsection III-C, the precision of speech transcription based on the data self-disclosure from *Amazon* is examined. subsection III-D investigates the speech recognition of the devices based on two different voice profiles (male and female) and a varying distance to the output speaker, before the output sound pressure of the individual *Echo* devices are finally compared with each other in subsection III-E.



(a) Setup for testing all models simultaneously.

(b) Setup for testing the individual devices with variable distance to the output speaker.

Fig. 1: Setups of the test environment.

For the investigations of subsection III-B and subsection III-C all *Echo* devices are placed in a semicircle in front of two output speakers, so that the distance between the *Echo* devices and the output speaker was almost identical (see Figure 1a). We then played out test sequences, listed in Table II, over the output speaker for 24 hours. This sequences was previously generated with three different voices profiles (2x Female; 1x Male) using the *Any Text to Voice*[1] service from *AnywaySoft Inc.*. In order to minimize resonances, the devices were placed on a sound-absorbing mat and the room was additionally equipped with sound-absorbing foam.

[1]https://www.anywaysoft.com/text-to-voice/download.html

name	time	original command	translation
volume	hh:01	Alexa, stelle die Lautstärke auf X. *(X changed hourly)*	Alexa, set the volume to X *(X changed hourly)*
news	hh:06	Alexa, was sind die Nachrichten?	Alexa, what's the news?
appointment	hh:11	Alexa, wann findet mein nächster Termin statt?	Alexa, when is my next appointment?
weather	hh:16	Alexa, wie wird das Wetter heute?	Alexa, what's the weather like today?
lotto	hh:21	Alexa, sag mir die Lottozahlen.	Alexa, tell me the lotto numbers.
shopping list	hh:26	Alexa, was steht auf meiner Einkaufsliste?	Alexa, what's on my shopping list?
foreign word	hh:31	Alexa, was bedeutet kongruent?	Alexa, what does congruent mean?
capital	hh:36	Alexa, was ist die Hauptstadt von Österreicht?	Alexa, what is the capital of Austria?
DAX	hh:41	Alexa, wie steht der DAX?	Alexa, how's the DAX?
skill	hh:46	Alexa, öffne Glückskeks.	Alexa, open fortune cookie.
sun	hh:51	Alexa, wann geht die Sonne auf?	Alexa, when does the sun come up?
football	hh:56	Alexa, auf welchen Tabellenplatz steht Dortmund?	Alexa, which place is Dortmund in the table?

TABLE II: Commands played during a test period of 24 hours including translation.
The sequences was repeated every hour, where the column *time* indicates when the command was played within an hour. The output voices profiles used also changed randomly every hour.

A. Technical Specifications

Table III shows the differences between the individual *Echo* models based on their specification.
The models differ especially in the existence of display, camera and ZigBee-hub and the number of microphones and speakers. This technical information can influence the quality of speech recognition, the output sound pressure and the power consumption of the devices.

	Echo Show	Echo Show 8	Echo Show 5	Echo (2. Gen)	Echo (3. Gen)	Echo Plus	Echo Flex	Echo Dot	Echo Studio
display	10.1"	8"	5.5"	None	None	None	None	None	None
camera	5MP	1MP	1MP	None	None	None	None	None	None
number speaker	2	2	1	2	2	2	1	1	5
microphone characteristic	far field	far field	far field	far field	far field	far field	far field	far field	far field
number microphones	4	2	2	7	7	7	2	4	7
ZigBee	yes	no	no	no	no	yes	no	no	yes
firmware (per 20/04/2020)	651614220	3725204100	3725204100	653621620	3658075268	3658075268	3658075268	3658075268	3725203076

TABLE III: Technical specification of each model according to Amazon product website.

B. Power Consumption

During the 24 hour test period, we recorded the power consumption of the individual *Echo* device by using *ALL3075v3* single power measure tools from *ALLNET GmbH*, Germering. These devices allows measurements in a resolution of up to 7Hz, which can be accessed via a local network interface (API). We stored and resampled the data to a frequency of 1Hz and then calculated the average value and the maximum value for power consumption of the individual *Echo* device over the entire test period of 24 hours.
Further we analyzed the power consumption data during playing radio (station: 'Bayern 1', Volume: 8) and in standby mode. We also investigated a period of 24 hours in each case. This values are listed in Table IV.

power consumption	Echo Show	Echo Show 8	Echo Show 5	Echo (2. Gen)	Echo (3. Gen)	Echo Plus	Echo Flex	Echo Dot	Echo Studio
test sequences (avg)	6.7	4.7	3.6	2.1	2.6	3.0	1.7	1.7	4.1
test sequences (max)	13.5	8.6	6.1	6.7	13.8	15.7	2.8	7.5	35.2
play radio	11.4	8.4	5.1	2.1	4.9	5.6	2.1	2.6	6.5
standby	9.3	6.3	4.3	2.8	3.5	4.15	2.0	2.2	6.2

TABLE IV: Average (avg) and maximum (max) power consumption during playing the test sequences, while playing radio and in standby mode over a test period of 24 hours per model (in *W*).

Even if all *Echo* devices received the same commands at the same time, some test sequences could not be understood or processed (see subsection III-C). This can have a direct impact on the power consumption.

In general, it can be observed that the *Echo Show* has the highest average power consumption (6.7W), followed by the *Echo Show 8* (4.7W) as well by playing the test sequences, playing radio or in standby mode. This can be explained by the displays of these devices, which consumes constant power. The highest peak consumption was received by the *Echo Studio* (35.2W) which corresponds to the higher number of included speakers and the integrated 330W amplifier.

C. Speech Transcription

To check how accurately speech recognition was performed during the 24 hour test phase, a data self-disclosure from Amazon was requested and analyzed for all devices. The data self-disclosure lists all commands understood by the device and contains transcriptions of the spoken phrases or 'Data not Available' entries. 'Data not Available' indicates that the activation word ('Alexa') was understood but not the command. Further the transcripted entries must be checked whether the command was interpreted correctly - i.e., if the device transcribed the spoken text correctly.
The transcriptions were analyzed in detail and compared with the expected values (see Table II). For the individual *Echo* devices the number of correct entries is shown in Table V.

command	Echo Show	Echo Show 8	Echo Show 5	Echo (2. Gen)	Echo (3. Gen)	Echo Plus	Echo Flex	Echo Dot	Echo Studio
volume	12	20	16	13	18	16	17	13	15
news	15	18	17	18	19	14	12	18	10
appointment	11	13	11	12	9	12	11	10	13
weather	16	17	13	12	17	17	16	14	17
lotto	12	18	15	17	17	18	12	15	12
shopping list	13	20	15	14	18	14	15	18	12
foreign word	8	7	3	9	6	11	4	9	3
capital	13	13	19	13	14	11	16	13	17
DAX	14	17	18	12	16	14	14	17	14
skill	18	18	17	19	20	13	16	16	19
sun	15	17	15	11	16	11	11	12	16
football	7	12	13	8	11	14	19	10	9
SUM	154	190	172	158	181	165	163	165	157

TABLE V: Number of correctly understood commands during testing all devices simultaneously for 24 hours according to the data self-disclosure. Maximum possible per command: 24 / sum: 288.

None of the tested *Echo* models was able to interpret all commands correct. The most reliable model was the *Echo Show 8*. It is particularly notable that the foreign word was often not or not correctly understood. The word 'kongruent' (transl. congruent) was often interpreted as 'Konkurrenz' (transl. competition).

D. Distance based Speech and Voice Recognition

In a second test setup, the *Echo* devices were tested individually. It was the intend to investigate how the distance of an audio source to the *Echo* device affect on the number of commands are understood or misunderstood. Therefore, the *Echo* devices and an output speaker was placed as shown in Figure 1b and the distances between them was varied. The *Echo* device was constantly standing on the front edge of a shelf - 35cm away from a wall and set to volume 5. The output speaker was set to 50dB[2]. We varied the distances and played an automated test sequence which was again created using the *Any Text to Voice* service from *AnywaySoft, Inc.*. The sequence asked ten times with a female voice and ten times with a male voice 'Alexa, wie viel ist eins plus ein' (transl.: Alexa, how much is one plus one). The number of correctly understood and answered commands was counted. The results are shown in Table VI.

distance *in cm*	voice profile	Echo Show	Echo Show 8	Echo Show 5	Echo (2. Gen)	Echo (3. Gen)	Echo Plus	Echo Flex	Echo Dot	Echo Studio
200	female	10	10	10	8	9	8	10	9	9
200	male	1	10	10	0	0	2	10	1	8
400	female	10	9	8	9	8	9	9	8	9
400	male	9	9	9	5	5	6	9	4	7
600	female	10	9	8	4	5	5	9	6	9
600	male	5	8	8	6	0	5	8	1	7
800	female	8	9	9	1	1	4	7	3	9
800	male	5	9	9	2	0	1	7	1	4
SUM	female	38	37	35	22	23	26	35	26	36
SUM	male	20	36	36	13	5	14	34	7	26
SUM	both	58	73	71	35	28	40	69	33	62

TABLE VI: Number of correlctly understood commands based on the distance between the *Echo* device and output speaker. Maximum possible per command: 10.

It can be generally stated that, if the output speakers are closer to the *Echo* device, the speech recognition will be more accurate. However, it is noteworthy that in some models there are significant differences in the precision of the male and

[2]DBA - measured 50cm in front of the device

female test sequences. On average, the male voice is recognized far less frequently than the female voice.
The *Echo Show 8* can again be identified as the most reliable model.

E. Output Volumes

In order to determine the individual output volume of each *Echo* device, the devices have again considered individually. The devices were placed on a sound absorbing mat and the room was equipped with sound absorbing foam (no anechoic chamber). The devices are then asked 'Alexa, wie viele Menschen leben auf der Erde?' (transl. Alexa, how many people live on earth?) and the sound pressure of their response was measured. For this purpose the measuring device 'DEM202' from the company 'Vellemann' was used, which was placed 100cm in front of the *Echo* device and 30cm above table height in the DBC-mode. This measuring device promises an accuracy of +/- 1.4dB.
The measurements are listed in Table VII.

volume Echo	Echo Show	Echo Show 8	Echo Show 5	Echo (2. Gen)	Echo (3. Gen)	Echo Plus	Echo Flex	Echo Dot	Echo Studio
1	51.2	54.8	58.3	40.8	55.2	54.6	38.0	41.5	51.4
2	50.7	56.9	58.6	39.5	59.8	59.6	43.8	55.4	53.2
3	58.0	62.4	64.7	55.9	65.1	64.9	46.3	61.7	58.9
4	61.5	68.3	70.9	63.3	71.2	71.4	46.7	67.2	64.7
5	67.0	71.3	71.4	67.5	73.8	73.8	50.6	70.0	67.4
6	69.6	74.0	72.0	72.3	74.9	75.6	51.6	74.5	70.1
7	75.2	77.2	74.0	76.9	78.3	78.1	53.8	75.8	72.0
8	78.3	78.4	74.8	82.4	81.3	81.3	58.1	76.1	76.1
9	80.6	84.5	74.9	83.2	83.0	82.6	58.7	77.0	78.2
10	82.5	84.4	76.0	83.8	82.7	82.4	59.1	76.7	78.1

TABLE VII: Output volume in dB (DBC) measured 100cm in front of the *Echo* device depending on the set volume.

IV. Conclusion And Future Work

The goal of this work was to investigate how the individual *Echo* models differ from each other in the aspects of speech recognition precision and output volume in a laboratory environment. Furthermore, we investigated how the power consumption of the models differs.

In two different test setups for measuring the speech recognition precision, it was observed that the model *Echo Show 8* correctly recognized the most commands by comparison to the other *Echo* devices.
In general it is worth mentioning that none of the tested models was able to interpret all commands correctly. When testing 288 sequences in maximum 190 (= 66%) was correctly understood.
For the models *Echo 2. Gen*, *Echo 3. Gen*, *Echo Plus* and *Echo Dot*, it was found that male voice sequences can not be recognized reliable. Similar can be observed for the models *Echo Show* and *Echo Studio*, but not in the significance. Specific problems interpreting the male voice did not occur with the *Echo Show 8*, *Echo Show 5* or *Echo Flex* models.
A correlation between the number of installed microphones and the number of correctly recognized commands could not be detected.

In investigations on the output sound pressure, we discovered that the model *Echo Show 8* reaches the highest output sound pressure (round 84dB). The models *Echo Show*, *Echo 2. Gen*, *Echo 3. Gen*, *Echo Plus* are each at about 83dB. The most silent model is *Echo Flex*. Conversely, however, it could not be observed that the *Echo Studio*, even if it had the highest number of integrated speakers, reached the highest sound pressure level. This can be attributed to the fact that this model is designed for 3D audio systems and the large number of speakers required to implement this.

With subsection III-B we have compared the average and maximum power consumption of the *Echo* models. We found that the display models consume on average more power than the non-display models (with the exception of the *Echo Studio*).

Our work showed that the voice profile of the person speaking with *Echo* is highly relevant for the reliability of speech recognition. In further work, the reliability of speech recognition with more different voice profiles (possibly also with dialect) has to be investigated. In addition, external factors such as background noise must be taken into account in further work.

References

Canalys. (2019). *Canalys: Amazon smart speaker shipments crossed 10 million mark in q3 2019.* Retrieved 3/6/2020, from `https://www.canalys.com/newsroom/worldwide-smartspeaker-Q3-2019`

Frawley, R. J. (2018). *Logging and analysis of internet of things (IoT) device network traffic and power consumption* (Doctoral dissertation). doi: 10.15368/theses.2018.60

Gao, Y., Pan, Z., Wang, H., & Chen, G. (2018, October). Alexa, my love: Analyzing reviews of amazon echo. In *2018 IEEE SmartWorld, ubiquitous intelligence & computing, advanced & trusted computing, scalable computing & communications,*

cloud & big data computing, internet of people and smart city innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI). IEEE. Retrieved from https://doi.org/10.1109/smartworld.2018.00094 doi: 10.1109/smartworld.2018.00094

Hellwig, A., Schneider, C., Meister, S., & Deiters, W. (2018). Sprachassistenten in der pflege - potentiale und voraussetzungen zur unterstützung von senioren. Retrieved from http://dl.gi.de/handle/20.500.12116/16665 doi: 10.18420/MUC2018-MCI-0341

Lloyd, C. (2018). How much electricity does the amazon echo use? *How-To Geek*. Retrieved 6/3/2020, from https://www.howtogeek.com/348219/how-much-electricity-does-the-amazon-echo-use/

Lopatovska, I., Rink, K., Knight, I., Raines, K., Cosenza, K., Williams, H., ... Martinez, A. (2018, March). Talk to me: Exploring user interactions with the amazon alexa. *Journal of Librarianship and Information Science*, *51*(4), 984–997. Retrieved from https://doi.org/10.1177/0961000618759414 doi: 10.1177/0961000618759414

Shalini, S., Levins, T., Robinson, E. L., Lane, K., Park, G., & Skubic, M. (2019). Development and comparison of customized voice-assistant systems for independent living older adults. In *Lecture notes in computer science* (pp. 464–479). Springer International Publishing. Retrieved from https://doi.org/10.1007/978-3-030-22015-0_36 doi: 10.1007/978-3-030-22015-0_36

Yamada, S., Kitakoshi, D., Yamashita, A., Suzuki, K., & Suzuki, M. (2018, November). Development of an intelligent dialogue agent with smart devices for older adults: A preliminary study. In *2018 conference on technologies and applications of artificial intelligence (TAAI)*. IEEE. Retrieved from https://doi.org/10.1109/taai.2018.00020 doi: 10.1109/taai.2018.00020