Conversation Frames: Yet Another Contextual Dimension for IPAs

Omar Saad Almousa
Jordan University of Science and Technology
Irbid, Jordan
osalmousa@just.edu.jo

Hazem Migdady
Oman College of Management and Technology
Muscat, Oman
hazem.migdady@omancollege.edu.om

***Abstract*—At first, we aimed at comparing the only two Intelligent Personal Assistants (IPAs) that support Arabic language, namely Siri of Apple and Salma of Mawdoo3. We compared them according to two features concerning the given answers: 1) Correctness and 2) Naturality. During our evaluation, we found out that both IPAs have a shortcoming that is present for other languages as well. This shortcoming is their inability to take into consideration the contextual information derived from consequent questions. To address this shortcoming, we propose a model that improves the naturality of an IPA without downgrading the correctness of its answers.**

I. Introduction

The foundation for Intelligent Personal Assistants (IPAs) as we know them today was the IBM Simon 1994 [1]. After that Siri of Apple was introduced as a novel feature acts as an IPA on iPhone 4s, precisely on Oct., 4th 2011 [2], [3].The authors of [4] present an overview of Siri as an IPA to illustrate the evolution in man-machine interaction. They mention the major features of the IPAs: 1- easiness of interaction 2- flexibility and 3- simplicity. Emphasizing that voice-based interaction is the easiest way for interaction because of its simplicity, flexibility and does not need cognitive efforts. However, IPAs still have some constraints in use such as the complexity in human speech and the varying context. That's why all available IPAs until now are used for specific description. However the industry of IPAs was accelerated and several IPAs were launched into the market.

The most common IPAs are Cortana of Microsoft [5], Alexa of Amazon [6], and Google Assistant [7]. In [8] a review of applying voice based personal assistants to bridge the gap of interaction between real and physical worlds is provided. Regarding Siri for instance, it applies the deep learning in AI that helps to monitor user activities to provide some personalized recommendations. The authors of [9] argue that AI along with machine learning has many areas to be propelled in, such as cognitive and learning sciences, game design, psychology, etc. Nowadays the market of IPAs witnesses a giant evolution since the value of an IPA can be expressed money wise. The authors of [10] and [11] agreed that the market volume of the IPAs industry will exceed the threshold of $2 billion by 2020. Moreover, the authors of [12] and [13] were even more optimistic about the predicted market share for the IPAs since it is estimated to reach $4.6 billion in less than a year from now. In addition [14] mentioned that the number of IPAs users may reach 1.8 billion in 2021 in an amount of increment more than $1.4 billion in comparison with the number of users in 2015. This would increase the revenues to be $15.8 billion instead of $1.6 billion in 2015. However the authors of [10] suggested some concerns about IPAs related to the matters of security and privacy. Such risks critically affect the reliability and trustworthiness of the IPAs. Hence the authors of [15] proposed an architecture to enhance the user guide interaction with the IPAs to deliver the right request and its appropriate response. Even though the IPAs were discussed in a wide range of research papers that covered, almost, all the aspects related to them, the way how IPAs support and deploy Arabic language interactions wasn't discussed properly.

The authors of [16] present a comprehensive review to provide a concrete basis for any future research related to IPAs. They believe that, despite the fact about the researches of IPAs are multidisciplinary, the core of IPAs is that they are nothing but smart machines that combine several techniques to sense and influence the environment. Hence this paper identified the functional principles and research domains that are promising for the future researches in this field. Taking into consideration that [16] introduced a review for any future research related to IPAs.

In this paper, we discuss two IPAs, Salma of Mawdoo3 and Siri of iPhone since they are the only known IPAs that support Arabic language conversations. We noticed that both IPAs had a common shortcoming related to their lack of handling the conversation context. Note that several researchers use the term context to express the physical/emotional/functional attributes of the surrounding environment including the users. More precisely, context is defined as any information used to describe the situation of any object relevant to the user-application interaction [16]. For example, the context may be referred to the light illumination level of a room [17], anticipated or current actions of users [18], [19], and [20].

In this research, by context we mean: the semantic frame of a conversation. Furthermore, the term context is used for the same purpose we have in this paper [21]. Therefore, we use the term conversation frame to avoid any ambiguity with another usage of the term context. Henceforth we use the term frame to refer to our definition of context. As mentioned before, we consider the only two IPAs that support the Arabic language for conversations.

We start this as a comparative study on the performance of Salama and Siri in the Arabic language. We performed the comparison according to two features: 1) Correctness and 2) Naturality of the IPAs answers. We adopted these concepts from [22]. For abbreviation, we call both features together: Properness. To perform our comparison, we designed a questionnaire of 9 weather-related questions being the common topic that both IPAs promote their ability to answer weather-related questions. We asked both IPAs the 9 questions, and then asked 12 participants to answer a 5-point paper-based Likert-scale questionnaire to evaluate the properness of each IPA. Then, we analyzed the questionnaire results, and we found a common limitation in both IPAs. The limitation is summarized by the inability to keep track of the conversation frame in subsequent questions. Finally, we propose a model to address this limitation, i.e., handle conversation frame through subsequent questions. The latest is the main contribution of this paper, this paper introduces two novel contributions: 1) we claim to be the first to introduce such a model that improves the properness on any IPA regardless the language, natural language processing modules, and deployed frameworks. 2) this research is the first one to evaluate IPAs that support the Arabic language that is the fifth world-wide language in number of speakers according to the Swedish encyclopedia with almost 300 million speakers in 2007 [23].

The rest of this paper is organized as follows: In Section II we give more details about our evaluated systems. The evaluation methods and results will be presented in Sections III and IV respectively. Section V introduces the proposed operational model to handle conversation frame, and we conclude in Section VI.

II. Evaluated Systems

In this section, we describe the evaluated IPAs in this paper. Siri is an intelligent personal assistant that was released in 2010 and dedicated to the iPhone devices only. It provides a variety of functions that controls the host device and provides users with a range of services. Similarly, Salma is an intelligent personal assistant that is designed to be an Arabic version of Siri of iPhone and Alexa of Amazon. Salma is currently redesigned voice-based Arabic interface service for businesses in various sectors such as telecommunication and electronics … etc. Table I below provides a summary about the IPAs under consideration.

TABLE I. Summary of IPAs

IPA	Manufacturer	Release Date	Supported Languages
Salma	Mawdoo3	2018	1 (Arabic)
Siri	Apple	2010	20 (Includes Arabic)

III. Evaluation Methods

In this section, we describe our evaluation methods. First, we noticed that both IPAs (Siri and Salma) promote their ability to answer weather-related questions. Being the only intersection between them, we decided to examine them based on it. To that end, we came up with 9 questions. The original Arabic questions are shown in Table II along with the English translation for each one of them.

TABLE II. Evaluation Questions in Arabic and Their English translation

Q1	ما هي حالة الطقس؟ What is the weather?
Q2	ما هي سرعة الرياح؟ What is the wind speed?
Q3	ما هي درجة الحرارة؟ What is the temperature?
Q4	ما هي نسبة الرطوبة؟ What is the humidity?
Q5	ما هي حالة الطقس في ابوظبي؟ What is the weather in Abu-Dhabi?
Q6	ما هي حالة الطقس غداً؟ (في أبوظبي) What is the weather tomorrow? (in Abu-Dhabi)
Q7	ما هي الملابس التي تنصحني بارتدائها؟ (في أبوظبي غداً) What clothes do you recommend me to wear? (in Abu-Dhabi tomorrow)
Q8	إقترح أماكن مناسبة للتنزه (في أبوظبي غداً) Suggest suitable places to visit. (in Abu-Dhabi tomorrow)
Q9	هل يمكنني السباحة الأسبوع القادم؟ (في أبوظبي) Could I swim next week? (in Abu-Dhabi)

As mentioned in section I, for our evaluation we use two quality concepts for a given answer, namely: Correctness and Naturality. An answer is correct if it is right, e.g., if the question is: “what is the capital of England?”, then “London” is a correct answer, while “Rome” is not. On the other hand, an answer is natural if it is delivered in a human-like fashion. For example, for the same question above, the answer “London” is natural, but the answer “London is located in England” is not. Naturality is more of a subjective (feeling) criterion that cannot be easily measured. Therefore, we later rely on a questionnaire by which different people measure the naturality of the IPAs covered in our study. In reference to Table II, note that to answer the first five questions properly (correctly and naturally), any assistant (machine or human) should answer a question according to the current time and location (current frame) unless it is explicitly given (uttered in the question). (Therefore, we later call the first five questions Direct Questions). Whereas for the remaining four questions (Q6 – Q9), the assistant should take into consideration some contextual information, e.g., to answer the sixth question properly, the assistant should consider that the location is Abu-Dhabi, since the previous question explicitly asks about the weather in Abu-Dhabi. Moreover, in the seventh question, the assistant should take into consideration that the time was changed in the previous question (Q6) and keep the location intact, i.e., Abu-Dhabi, since no explicit change for the location is mentioned after Q5.

Generally speaking, to answer any question properly an assistant should take into consideration a frame that is composed of three components: subject, location, and time. For example, the frame of question 1 is (weather, here, now), such that for subject=weather, location = here (the current location of the assistant host device), and time= now (the current time of the assistant host device). TableIII tracks the frame changes through questions Q1 - Q9. Later in Section IV, we present a concrete operational model to handle frame.

TABLE III. FRAME CHANGES

No.	Frame			Explicit/ Implicit
	Subject	Location	Time	
Q1	weather	here	now	Explicit Frame
Q2	wind-speed	here	now	Explicit Frame
Q3	temperature	here	now	Explicit Frame
Q4	humidity	here	now	Explicit Frame
Q5	weather	Abu-Dhabi	now	Explicit Frame
Q6	weather	Abu-Dhabi	tomorrow	Implicit Location (based on previous frame from Q5)
Q7	clothes	Abu-Dhabi	tomorrow	Implicit Location and Time(based on previous frame from Q6)
Q8	visiting-places	Abu-Dhabi	tomorrow	Implicit Location and Time(based on previous frame from Q7)
Q9	swimming	Abu-Dhabi	next-week	Implicit Location (based on previous frame from Q8)

According to Table III, we classify our questions into two categories: 1) Direct Questions (Q1-Q5) that have explicit frame changes caused by the current question itself. 2) Contextual Questions (Q6-Q9) that have implicit contextual changes caused by previous question(s). Hence, it is possible to say that: frame immigrates over questions unless an explicit change is enforced. After the preparation of our 9 questions, we started our experiment such that one of the researchers asked Siri and Salma to answer the questions in two different sessions using the same voice tone and pace. As a hosting device, we used iPhone 6s plus, and we recorded each session separately using iPhone 5s. Then, we uploaded the recorded videos on YouTube under the URLs shown in Table IV. (We have not used a URL shortener to preserve the privacy of our readers).

TABLE IV. URLS FOR RECORDED SESSIONS

Siri session	https://www.youtube.com/watch?v=aFHXUYe1a-w
Salma session	https://www.youtube.com/watch?v=FmGxyQjUeNo

Once we prepared the two videos, we asked 12 participants familiar with smartphones applications (average age=30, StaDev ± 8, females = males = 6) to answer a 5-point paper-based Likert-scale questionnaire about the correctness and naturality of the answers given by the tested IPAs to our 9 questions. We recruit the participants according to personal communications with colleagues, friends, and family. Two of the participants hold a PhD degree, and the rest hold a BSc/ BA degrees. We present the results of the evaluations in the next section.

IV. RESULTS

In this section, we present the results of the evaluation that we described above. As mentioned earlier, the evaluation method was a questionnaire that consists of 9 questions. Each of the questions has a 5-point Likert-scale. We asked each participant to watch the videos that were uploaded on YouTube (see Table V), and then fill the questionnaire accordingly. The participants were allowed to pause, rewind, and replay the videos as they wish in order to answer the questions conveniently.

TABLE V. QUESTIONNAIRE RESULTS

Salma-Correctness

No.	All	Females	Males
Q1	4.75	4.83	4.67
Q2	2.58	2.33	2.83
Q3	4.25	4.17	4.33
Q4	2.08	2.33	1.83
Q5	4.67	4.67	4.67
Direct AVG	3.67	3.67	3.67
Q6	3.17	3.50	2.83
Q7	2.08	2.50	1.67
Q8	1.58	1.83	1.33
Q9	2.08	2.33	1.83
Contextual AVG	2.23	2.54	1.92
Difference	1.44	1.13	1.75

Siri-Correctness

No.	All	Females	Males
Q1	4.67	4.67	4.67
Q2	4.67	4.67	4.67
Q3	4.67	4.83	4.5
Q4	4.42	4.67	4.17
Q5	4.17	4.17	4.17
Direct AVG	4.52	4.60	4.43
Q6	2.75	3.33	2.17
Q7	1.50	1.67	1.33
Q8	2.25	2.50	2.00
Q9	2.42	2.67	2.17
Contextual AVG	2.23	2.54	1.92
Difference	2.29	2.06	2.51

Salma-Naturality

No.	All	Females	Males
Q1	4.58	4.50	4.67
Q2	2.83	2.33	3.33
Q3	4.08	3.83	4.33
Q4	2.58	2.33	2.83
Q5	4.42	4.00	4.83
Direct AVG	3.70	3.40	4.00
Q6	3.58	3.67	3.50
Q7	2.50	2.33	2.67
Q8	1.83	1.33	2.33
Q9	2.83	3.00	2.67
Contextual AVG	2.69	2.58	2.79
Difference	1.01	0.82	1.21

Siri-Naturality

No.	All	Females	Males
Q1	4.25	4.33	4.17
Q2	4.17	4.33	4.00
Q3	3.50	4.00	3.00
Q4	3.50	4.17	2.83
Q5	3.25	3.33	3.17
Direct AVG	3.73	4.03	3.43
Q6	3.00	3.50	2.40
Q7	1.50	1.33	1.67
Q8	2.83	3.17	2.50
Q9	2.83	3.00	2.67
Contextual AVG	2.54	2.75	2.31
Difference	1.19	1.28	1.12

Henceforth we follow this order: First we present the results of correctness for the direct questions, then the results of correctness for the contextual questions. After that, we present the results of naturality in the same manner. Finally, we take the gender of each participant into consideration. Table V summarizes the results of the evaluation. The table illustrates the averages of the direct and contextual question answers separately. The last row shows the differences between the averages, which provide a better understanding of the performance of the IPAs with respect to correctness, naturality, and participants' gender. Note that we use the averaging of discrete Likert-scale score to indicate differences between different systems and groups, and we have not used them in any mathematical computations.

A. *Correctness Remarks*

For direct questions (Q1 – Q5), Siri scored better than Salma (4.52 to 3.67). However, in contextual questions, both assistants' scores dropped to reach a draw at 2.23. This implies that the amount of "loss" for Siri was greater than that one of Salma. In general, neither assistant detected the frame change in the contextual questions. Fig. 1 depicts the results on correctness.

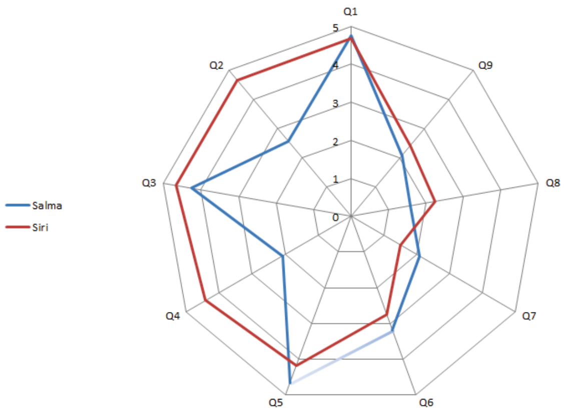

Fig. 1. Correctness Results

B. *Naturality Remarks*

For direct questions, Siri has a slight advantage over Salma (with 0.03). However, in the contextual questions, the amount of "loss" of Siri was greater than that one of Salma. An interesting remark here is that, with respect to naturality feature, the performance of Salma over contextual questions was better than that one for Siri, i.e., Siri scored 2.54 and Salma scored 2.69. Fig. 2 illustrates the naturality results.

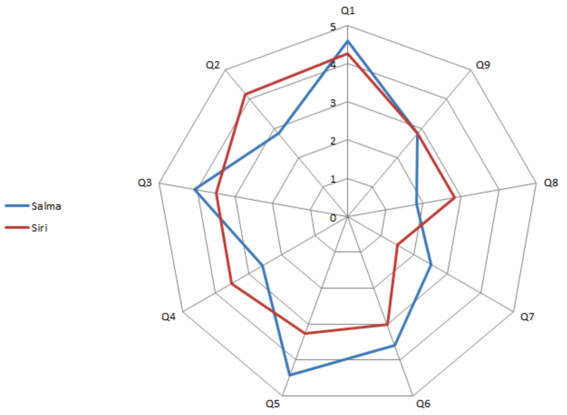

Fig. 2. Naturality Results

C. *Gender-based Remarks*

For the correctness of Salma answers for direct questions, we don't notice a correlation between the results and the participants' gender (the Direct AVG is 3.67). However, for contextual questions, males are less satisfied than females (Contextual AVG is 1.92 for males compared to 2.54 for females). On the other hand, the correctness of Siri answers for both types of questions is more satisfying for females than for males. Fig. 3 shows the gender-based results for Salma's correctness.

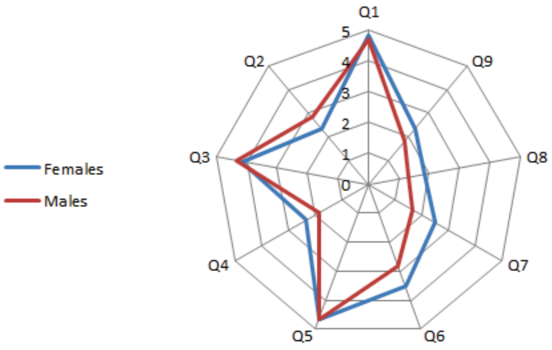

Fig. 3. Gender-based Results for Salma Correctness

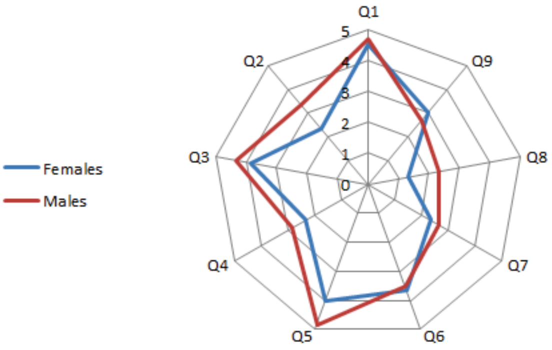

Fig. 4. Gender-based Results for Salma Naturality

In comparing the two IPAs, the amount of Siri's "loss" in the correctness of contextual questions is greater than that one of Salma, i.e., the average score for the correctness of Siri for direct questions is: 4.60 for females, and 4.43 for males, while the average score for the correctness of Siri in contextual questions is: female=2.54, males=1.29 with amount of "loss" 2.06 for females, and 2.51 for males. Whereas, the amount of loss for Salma is: 1.13 for females and 1.75 for males. Thus, the deviation of Salma is less than that one of Siri. Fig. 4 shows gender-based results for the naturality of Salma. While Fig. 5 and Fig. 6. show the gender-based results for the correctness of and naturality of Siri respectively.

In general, the assistants are able to satisfy females more than males except in the case of Salma naturality, i.e., for direct questions females score is 3.40 compared to 4.00 for males, and in contextual questions, females score is 2.58 compared to 2.79 for males. Despite that, the fall of males' satisfaction after contextual questions is greater than that one for females

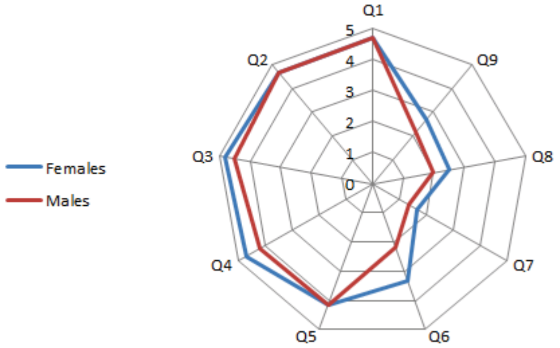

Fig. 5. Gender-based Results for Siri Correctness

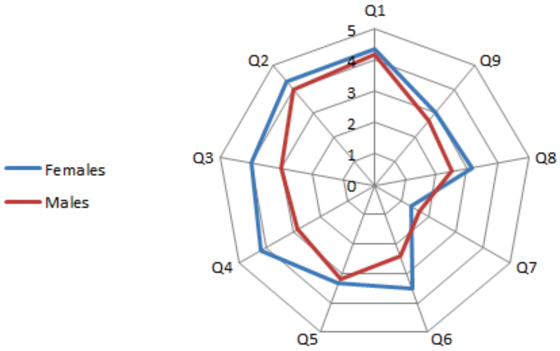

Fig. 6. Gender-based Results for Siri Naturality

(1.21 to 0.82). It is crucial to emphasize that: the failure to detect frame change has negatively affected the performance of both IPAs.

V. Frame Modelling

The major result we notice from the previous section is that: there is a shortcoming in frame handling in both IPAs, i.e., IPAs fail to keep track of conversation frame through subsequent questions. This is shown by the fact that the Contextual AVG is down below the Direct AVG (cf. Table V). We also noticed that: this shortcoming is partially present in the English version of Siri as Salma has only an Arabic version.

Therefore, we propose a model to improve the answer-properness for any IPA (answer correctness and naturality). To that end, we use the term conversation to call a sequence of questions from the user of an IPA and answers/responses from the IPA to the given question. For any IPA, to improve its properness when answering a question, the IPA should maintain a conversation frame that keeps track of the information flow through questions.

A conversation frame (henceforth a frame) should contain information about the subject, location, and time. More precisely, a frame Fi is a tuple (Si, Li, Ti) where Si is the subject of Fi, Li is the location of Fi, and Ti is the time of Fi. For formalities, one can refer to them as subject(Fi), location(Fi), and time(Fi) respectively. We extend these projectors to questions and say subject(Qj) to denote the subject of the jth question Qj, and the same applies for location(Qj), and time(Qj). In case that a certain question Qj has no mentioning for a specific subject, location, or time, then we use the generic value NULL (with some data-type abuse with an apology to the type-restricted audience). For example, let Qj= “How is the weather?”, then subject(Qj) = weather, and location(Qj)=NULL, and time(Qj) = NULL. To answer a question (say Qj to denote the jth question) properly (correctly and naturality), an IPA should maintain two frames:

1) a pre-frame denoted by Fj and represents the frame before questioning Qj, i.e., before the user utters her/his question.
2) a post-frame denoted by ϕ_j and represents the frame after questioning Qj, i.e., after the user utters her/his question.

In general, we define a global initial frame F1 = (USER, HERE, NOW), where USER is the user of the IPA hosting device, HERE is the current location, and NOW is the current time according to the host device. Moreover, ϕ_j=Fj unless Qj has a subject, location, or time different than subject(Fj), location(Fj), and time(Fj) respectively. More precisely, we have subject(ϕ_j)= subject(Qj), location(ϕ_j)=location(Qj), and time(ϕ_j)=time(Qj). Now we come to the operational part of frame update in a conversation. For any given question Qj, we have:

- Fj=ϕ_j-1
- if subject(Qj) = NULL then subject(ϕ_j) = subject(Fj), otherwise subject(ϕ_j) = subject(Qj)
- if location(Qj) = NULL then location(ϕ_j) = location(Fj), otherwise location(ϕ_j) = location(Qj)
- if time(Qj) = NULL then time(ϕ_j) = time(Fj), otherwise time(ϕ_j) = time(Qj)

In this way, we update the frame only if there is an update for any of its parts in the question utterance; otherwise, we keep the frame as is (the same as the previous question). As an example, let us run our model on the conversation (sequence of questions) listed earlier in Table II and keep track of have an example in which we apply.

j	Fj (pre-frame)	ϕ_j (post-frame)
1	(USER, HERE, NOW)	(weather, HERE, NOW)
2	(weather, HERE, NOW)	(wind-speed, HERE, NOW)
3	(wind-speed, HERE, NOW)	(temperature, HERE, NOW)
4	(temperature, HERE, NOW)	(humidity, HERE, NOW)
5	(humidity, HERE, NOW)	(weather, Abu-Dhabi, NOW)
6	(weather, Abu-Dhabi, NOW)	(weather, Abu-Dhabi, tomorrow)
7	(weather, Abu-Dhabi, tomorrow)	(clothes, Abu-Dhabi, tomorrow)
8	(clothes, Abu-Dhabi, tomorrow)	(visiting-places, Abu-Dhabi, tomorrow)
9	(visiting-places, Abu-Dhabi, tomorrow)	(swimming, Abu-Dhabi, next-week)

In implementation, the two frames can be handled with a single object that is updated each time a question is asked by a user and before its getting answered by an IPA. As explained above, the pre-frame is the result of answering a sequence of previous questions. In the case of the very first question we have F1 (the initial frame). Implementation wise, a conversation can be bounded by the start and end of a question/answer session. Assume a person p wants to ask her/his assistant about n different features: f1,..., fn concerning a location l on a certain date d such that: l and d are not the current location and date, then p needs to ask n questions of the form:

Q1: What is f1 of l on d?

Q2: What is f2 of l on d?

. ...

. ...

Qn: What is fn of l on d?

Notice the amount of redundancy in questions Q2,...,Qn (denoted by underline). Note that in a natural conversation, such redundancy is neither needed nor comfortable. Note that adding "auxiliary" words such as "there" and "that time" will help the assistant to maintain some sort of context, but this is neither comfortable nor natural.

In our model, we avoid such redundancy by preserving a frame for each conversation. In our model, p needs to ask Q1 as is, but for the rest of the conversation: questions Q2, ..., Qn will be reduced by omitting the underlined part of each of them (our model keeps track of l and d as long as no changes occur to cause otherwise). Moreover, our model is very general and covers a vast range of conversations since subject, location, and time features are essential in almost any conversation. Hence, we do claim that: our model improves the naturality of the conversation without downgrading the correctness of the answers.

Finally, this model is generalizable to other features for different types of conversations. For example, some possible features to extend our model are the intent to capture the purpose of the question such as requesting information or issuing an order. Another feature may be the sentiment i.e., the emotional attitude of the user. A simple yet generic implementation of such model can be achieved by preserving a conversation frame in the form of a list of attribute-value pairs.

An alternative to our model could be a model that enables the IPA to respond with a clarification question. So in our case, if the user asks about the weather, the IPA can answer by a question about the time. After the user answers the time question, then the IPA can issue another question about the location. For example, let us have this conversation:

USER What is the weather?

IPA In which time you want me to answer?

USER now

IPA In which location you want me to answer?

USER here.

Of course this example is some how extreme, but this only to show how bizarre it would be to deploy a model that does not have any initial frame values and that does not follow any conversation frame. However, this method of responding to a question by a question may be a solution in case of what we refer to by blur questions. A blue question is that question that is ambiguous and can not be answered normally unless some of its ambiguity is clarified. For example, a question like "what's up?" are blur questions. In such case, we suggest that the IPA answers with an option of two: 1) An IPA can use a standard general answer such as "Not too bad!" as exemplary answers to the exemplary blur questions above respectively. 2) An IPA replies with a question to clarify the given blur question. In this case we may need an interrupting frame structure to handle such a situation. (interrupts and frame-switching mechanisms may inspire a solution). Worthy mentioning that the tested IPAs answer such questions by stating their failure to understand the given blur question. Before concluding, and aside from conversation frames, we want to point out some remarks about the two IPAs:

- Both IPAs needs improvement on language basis, i.e., the authors have notices several problems is cases, phonetics, and diacritics. In many cases such issues are tolerable, but this is not the general case. However, this venue is out of the scope of this research.
- In certain cases, both IPAs respond with showing (not uttering) a result of a web search for the given question. In this case, Salma is restricted to give an answer from mawdoo3.com website, while Siri conducts a general web search.

VI. Conclusion

Siri of Apple and Salma of Mawdoo3 are the only two IPAs that support the Arabic language. More precisely, Salma supports only the Arabic language. While Siri supports several languages including Arabic, and English. We started this study as a comparative study on the performance of the two IPAs in Arabic language. We selected two features for this comparison: correctness and naturality of the given answers to measure the properness of a given answer.

To perform our comparison, we designed 9 questions based on the weather being the common topic that both IPAs promote their ability to answer weather-related questions. We asked the IPAs the 9 questions, and we recorded a questioning answering session for each IPA. We uploaded the recorded videos on YouTube and asked 12 participants to answer a 5-point paper-based Likert-scale questionnaire to evaluate the properness of each IPA. We analyzed the questionnaire results, and we found out a common limitation in both IPAs. The limitation is summarized by their inability to keep track of the conversation frame in subsequent questions. Finally, we propose a model to address this limitation, i.e., handle frame through subsequent questions.

The next step of this research is to implement the proposed frame-handling model in order to compare it with existing IPAs. An implementation will facilitate a quantitative measurement of the effect of our model on the properness of IPA performance. We expect that such a study will prove a significant improvement in the properness of IPAs without neglecting the possibility of finding some limitations in our model that would be interesting to address. One can use existing tools/frameworks like Sphinx of CMU to carry out such implementation. Another potential future work can be extending this comparative study to cover more tools and languages.

An interesting extension of this work is testing other IPAs in terms of correctness and naturality. Adding more questions, asking more participants, and using better statistical tools and measures to demonstrate reliable differences among IPAs and users' groups could be also another venue for future work. Finally, a linguistic study will be very crucial to address evaluate such systems from linguistic point of view.

ACKNOWLEDGMENT

The authors thank the participants who answered the questionnaire.

REFERENCES

[1] Engadget website, "iPhone 4S Hands-on!," Web: www.engadget.com/2011/10/04/iphone-4s-hands-on/, Last accessed July, 17, 2019.

[2] The Official website of Apple, Apple Siri, Web: http://www.apple.com/ios/siri/, Last accessed 17-7-2019.

[3] C. Velazco, "Apple Reveals Siri Voice Interface: The "Intelligent Assistant" Only For iPhone 4S," TechCrunch, AOL.

[4] B. Shakeel, Tabasum, and M. S. Ahmad, "Siri – Apple's Personal Assistant : a Review," *International Journal of Computer Science and Mobile Computing (IJCSMC)*, vol. 6, issue 7, pg.44 – 48, 2017.

[5] Amazon website, Alexa Voice Service Overview (v20160207), Web: https://developer.amazon.com/docs/alexa-voice-service/api-overview.html, Last accessed June 20, 2019

[6] CNN website, "Growing Up with Alexa," Web: https://edition.cnn.com/2018/10/16/tech/alexa-child-development/index.html Last Accessed June 20, 2019.

[7] C. de Looper, "Google Wants to Make its Next Personal Assistant More Personable by Giving it a Childhood," Digital Trends, Last accessed July, 17, 2019.

[8] S. Madalli, R. Manikandan, V. Pandita, N. Surwade, and B. Shirgapur, "A Review of Voice Based Personal Assistants," *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, vol. 6, issue 2, pp. 1444-1447, 2018.

[9] V. Zue, S. Seneff, J. R. Glass, J. Polifroni, C. Pao, T. J. Hazen, and L.Hetherington, "JUPITER: A Telephone Based Conversational Interface for Weather Information," in *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 1, pp. 85–96, 2000.

[10] H. Chung, M. Iorga, J. Voas, and S. Lee, "Alexa, Can I Trust You?," *Computer*, vol. 50, issue 9, pp/ 100-104, 2017.

[11] Gartner website. "Worldwide Spending on VPA-Enabled Wireless Speakers Will Top $2 Billion by 2020" press release, 2017, Web: www.gartner.com/newsroom/id/3464317.

[12] B. R.Cowan, N. Pantidi, D. Coyle, K. Morrissey, P. Clarke, S. Al-Shehri, D. Early, and N. Bandeira, "What Can I Help You with?: Infrequent Users' Experiences of Intelligent Personal Assistants," in *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services, September, 2017*, ACM, 2017. pp. 43.

[13] Kamitis, Intelligent Personal Assistant- Products, Technologies and Market: 2017-2022, 2016.

[14] Tracitca website, "The Virtual Digital Assistant Market Will Reach $15.8 Billion Worldwide by 2021", Web: https://www.tractica.com/newsroom/press-releases/the-virtual-digital-assistantmarket-will-reach-15-8-billion-worldwide-by-2021/

[15] I. Hwang, J. Jung, J. Kim, Y. Shin and J. Seol, "Architecture for Automatic Generation of User Interaction Guides with Intelligent Assistant," in *Proceddings of the 31st International Conference on Advanced Information Networking and Applications Workshops (WAINA), Taipei, 2017*, pp. 352-355, 2017.

[16] R. Knote, A. Janson, L. Eigenbrod, and M. Söllner, "The What and How of Smart Personal Assistants: Principles and Application Domains for IS Research,“ in *Proceedings of Multikonferenz Wirtschaftsinformatik Conference, March, 2018*, Lüneburg, Germany, pp. 1083-1094, 2018.

[17] B. Ospan, N. Khan, J. Augusto, M. Quinde, and K. Nurgaliyev, "Context Aware Virtual Assistant with Case-based Conflict Resolution in Multi-user Smart Home Environment," in *2018 International Conference on Computing and Network Communications (CoCoNet), August, 2018*, pp. 36-44. IEEE, 2018.

[18] B. Ospan, "Simulation of a Simple Bio-Mimetic Robot with Neuromorphic Control System and Optimization Based on the Genetic Algorithm, *International Journal of Innovations in Engineering and Technology*, vol. 8, issue. 4, 2017.

[19] S. K. Das, N. Roy, A. Roy, "Context-aware Resource Management in Multi-inhabitant Smart Homes: A Framework Based on Nash H-learning," *Pervasive and Mobile Computing*, vol. 2(4), pp. 372-404, 2006.

[20] B. El Saghir, and N. Crespi, "An Intelligent Assistant for Context-aware Adaptation of Personal Communications," in *Proceedings of IEEE Wireless Communications and Networking Conference, 2007*, IEEE, 2007.

[21] E. C. Paraiso, and A. Jean-Paul, "An Intelligent Speech Interface for Personal Assistants in R&D Projects," *Expert Systems with Applications*, vol. 31, issue 4, pp. 673-683, 2006.

[22] G. López, L. Quesada, and L.A. Guerrero, "Alexa vs. Siri vs. Cortana vs. Google Assistant: a Comparison of Speech-based Natural User Interfaces," in *Proceeding of International Conference on Applied Human Factors and Ergonomics, 2017*, Springer, 2017.

[23] Ethnologue website, Web: https://www.ethnologue.com/language/ara, Last accessed August, 8, 2019 .