1 INTRODUCTION

~~Over more than a decade, r~~Research from over more than a decade has increasingly shown that humans as social beings imitate each other during conversations. This convergence occurs at many different level especially in linguistic behaviour~~,~~ and ~~it~~ may lead to easy and successful communication as argued by Pickering and Garrod [1][2]. The word *alignment* has been used to refer to such convergences.

~~Major~~ A major part of alignment studies cover~~s~~ dialogues between human interlocutors. Yet, as dialogues between humans and artificial devices (such as smart mobiles, computers, tablet devices~~-~~) ~~have~~ increasingly ~~becoming~~ become part of everyday life, investigating the characteristics of such dialogues will help us ~~to~~ understand how this interaction works and how it will help the development of more effective systems [3].

Additionally, an ~~existed~~ existent evidence of people treating computers as social actors and evaluating them as they evaluate people [4] strongly suggests the presence of alignment in Human~~—-~~Computer dialogue. Further~~more~~, Suzuki et al. (2003) report findings on the effects of voice prosody in human-computer interaction and show how the prosodic imitation by a computer of the user's voice increases the user's friendliness to the computer [5], ~~Which~~ which by potentially using the voice prosody alignment in spoken dialogue systems tries to obtain the user's emotional engagement [6].

To gain a better understanding of alignment, we have to identify its different aspects and recognise which are the mechanisms that ~~underlie~~ underlie it.

1.1 Levels of Alignment

There are many aspects of alignment involved in the communication process. To mention a few, alignment also occurs at a non-linguistic level ~~such as~~and includes facial expressions and emotional states. ~~but~~ But extensive research and studies have discovered that such alignment is widely spread at the linguistic level [3].

Interlocutors show a strong tendency to repeat each other's term of their lexical or grammatical choices [7–9]. For example, when two interlocutors both of who ~~both~~ use the same name for an object are said to be aligned on the use of that term even ~~of~~ if one or both of them never usually use it~~, then they are aligned on the use of that term~~. Similarly, with ~~in~~ respect ~~of~~ to syntactic representation, according to [7] naïve participants followed their confederate scripted description when he used a "prepositional object" (PO) form like *the doctor selling the gun to the artist* and produced another prepositional object description when it ~~is~~ was their turn to describe. ~~And t~~They did the same when the confederate produced a "~~-~~double object" ~~"~~ (DO) form description like *the doctor selling the artist the gun.*

Comment [Editor1]: Do you mean to say 'trie to emotionally engage the user'?

Comment [Editor2]: Should this be 'repeat ea other's lexical or grammatical choices'?

Comment [Editor3]: Is this space intentional? Please mae it consistent if so.

Comment [Editor4]: Describe what? Or do yo mean to say describe an event similarly?

Another level of alignment is at the level of the structure related to the meaning where different semantic representations can be correlated to an interlocutor's choice of structure. For example, when describing a spatial position, as speaker may use the same form of description and its interpretation. When a pair of participants played a cooperative maze game in a Garrod and Anderson (1987) experiment, they aligned their representations when the form *I'm two along, four up* was used and aligned again with the other form of description *I'm at B4* [3].

1.2 Alignment Mechanisms

Although it is unarguably to say that alignment between interlocutors occurs at many different levels and many proposed mechanisms that account for it, but our focus, in this project on linguistic level of alignment.

Comment [Editor5]: Unarguably correct/accurate or inaccurate? Please check for clarity

Comment [Editor6]: Please check for an incomplete sentence

1.2.1 Unmediated mechanisms of alignment

Priming The priming of representations is one of the proposed approaches to explain alignment and it is largely based upon this mechanism [1][3]. This priming is not mediated or influenced by prior beliefs about the addressee. Therefore, it called *unmediated* mechanism.

Studies of language processing suggested that the activation of such representations does not fade immediately and continues to increase activation increase of the possibility of using the same linguistic form by the speaker (e.g. [10]).

Comment [Editor7]: Please check for clarity

This unmediated alignment is consistent with evidence that interlocutors usually unaware of alignment have taken place [1]. Participants in the Branigan et al. (2011) experiment showed a very strong tendency to use the same choice of lexical words even if it was one of the disfavoured names specified in the experiment. For example, if their partner used the name *toad* instead of *frog* which was tested and classified as a disfavoured name for this creature, the naïve participants would ignore their default preferences and use the same disfavoured name that they had been exposed to [8].

Comment [Editor8]: Please check for clarity

1.2.2 Mediated mechanisms of alignment

In contrast with the previous mechanism, some speakers would may intentionally use the same linguistic representation with a particular interlocutor because they believed that this will would lead to a successful communication, or this is would be the appropriate expression to use, so their partners could understand them [3]. This alignment is mediated by the speakers' beliefs about their interlocutors and it is usually referred to as *audience design.*

A clear example of this mechanism, is evident when a speakers communicates with a non-native partner and they avoid using complex expressions or rare words or even speak more

slowly than when communicating with another native speaker~~,~~ and this effect highlights the fact that speakers form judgments about their partners' community membership and their linguistic knowledge. Consequently, such judgments influence their choice of linguistic representations [3].

Moreover, the alignment occurs ~~upon~~ based on the speakers' beliefs about their interlocutors' knowledge of the language ~~knowledge~~ based on their own experience with that particular interlocutor. For example, if their partner ~~used~~ uses a specific name for an object, the speaker may assume that this is a good word to use because their partner understands it. Similarly, if the speaker ~~used~~ uses the name *pram* for example, instead of *pushchair* and the interlocutor appears to understand it, then the speaker makes reference ~~of~~ to it and thinks it is a more suitable word to use [8].

1.3 Related Work

As mentioned earlier, this study will focus on the aspects of ~~our focus in this study is the~~ Human-Computer Interaction ~~aspects of this phenomenon~~ as many recent research in this area have investigated the effects and existence of alignment ~~in~~ between human-computer in written and spoken dialogues [3], [6–8], [11], [12]. Most of these researches employ~~ed~~ the same methods used to investigate the Human-Human alignment using a modified version of [7] confederate-scripting paradigm. Participants in such an experiment were told that they were playing a picture matching and description game with another partner in a different room connected by a network terminal. Although they believed that they ~~ware~~ ere interacting either with a computer or a human, ~~but~~ in fact they were always ~~interacted~~ interacting with a computer program that produced scripted utterances.

One of the interesting findings in [8] showed that ~~have shown that~~ lexical alignment with computers ~~is~~ was more stronger and robust compared to that with human and it ~~is~~ was largely mediated by the participants' beliefs about their interlocutors. Where the strength of linguistic alignment increased when they believed that they were interacting with a computer more than if they believed their partner is a person and this alignment is largely comparable. Additionally, participants displayed ~~showed~~ more alignment with a computer which appeared to them as being unsophisticated and old-fashioned compared to ~~than with~~ one that appeared ~~to be~~as being state-of-the-art and technically advanced.

However, [12] investigated the syntactic aspects of the linguistic alignment with computers. Their result supported the claim of Reeves and Nass's in [4] that people treated computers as social actors and added a linguistic domain into their work. Whereas findings about syntactic alignment in this study keeps up the claim that it is largely an automatic process and unmediated by beliefs about the interlocutor akin to Brennan and Clark's (1996).

Comment [Editor9]: Please be consistent abo spacing between text and the reference citation

Comment [Editor10]: Please check for clarity

Comment [Editor11]: Please check for clarity and incomplete sentenec

Although [8] analysis ~~provided~~ provides no evidence that lexical alignment is affected by written or spoken utterance~~,~~s, [5] have discovered that the voice prosody ~~effect~~ affects~~the~~ human-computer interaction in both behavioural and psychological terms. Thus, [6] also demonstrated a mutual alignment at the prosodic level and users' evaluation for the computer familiarity increased when the computer imitated their intonation. Therefore, they suggested ~~that~~ a mutual prosodic alignment in the human-computer interaction, which could be induced by the prosodic changes in a computer voice.

While most of the studies on alignment in HCI investigate~~d~~ the written input and lexical alignments, researchers in [11] ~~took~~ take the initiative to study syntactic alignments in spoken dialogue interactions within HCI and examine how the user beliefs about the system could be affected by choices of interface design ~~choices~~ (-such as the naturalness of the type of voice used by the interlocutor-) and the implication of this perception about the system on alignment in human-computer dialogues. Although the result of their research ~~suggested~~ suggests that the effect of voice type on syntactic alignment is not significant, they ~~found~~ find that participants' evaluation of the system in terms of interaction satisfaction ~~was~~ is significantly different between the experiment's three conditions (human , basic voice , advanced voice) but this is not ~~wasn't~~ reflected in the amount of alignment seen between the conditions ~~in~~ contrary ~~with~~ to the previous predictions. In contrast with [12] that suggests the existence of more mediated components in syntactic alignment between the humans and the computers, more- research ~~was~~ is proposed to investigate the scarcity of syntactic alignment in [11].

1.4 Aim and Objectives

One of the main reasons that explains ~~to explain~~ the lack of alignment differences between the experiment conditions in [11] is the syntactic structure used in the experiment and how the increased ~~how is the high~~ natural preference to use one structure (AN : Adjective Noun e.g. *red circle*) over another (RC : Relative Clause e.g. *circle that's red*) m~~ight have~~ay overrule~~d~~ any priming effect and ~~may~~ saturate~~d~~ any alignment effect. Therefore, it is necessary to explore whether ~~this~~ these findings ~~is~~ are also repeated with different syntactic structures though with less natural preferences such as the (PO) prepositional object and the (DO) direct object ~~that~~ used in [7], [10], [13] and ~~that~~ what we aim to investigate in this project. Whereas ~~in this study~~ our main objective in this study will only be ~~only~~ investigating the effect of the interlocutors' voice type~~s~~ on the syntactic alignment using (PO) and (DO) syntactic structure.

Comment [Editor12]: Please check for incomplete sentenec

1.5 Achievements

We employed the same experiment design that is usually used ~~to~~ in syntactic alignment research, specifically ~~that~~ what is used in [11] but with the (PO) and (DO) structure instead. Similarly, ~~We~~ we found that there was no significant effect of voice type on syntactic alignment ($p > 0.05$) but there was a significant effect of the type of prime on the alignment and that supported the claim of syntactic alignment ~~is~~ being more automatic. This ~~and~~ unmediated my beliefs about interlocutors.

Finally, from the previous observations on alignment, we have seen that it involves many different components at many different levels. However, recent studies have introduced effective methods to address ~~which~~ those components which are most relevant to alignment in HCI as ~~we are going to~~ discussed in the next chapter.

Comment [Editor13]: Please check for chang in content

2 METHODOLOGY

The main difficulty faced by the researcher ~~that may face the researcher to~~when ~~compare~~ comparing the alignment in HCI with that ~~alignment~~ between humans ~~is~~ may be that there is more than one factor that ~~could~~ can influence the participant's behaviour. One of the effective methods to address this matter is to use the Wizard of Oz paradigm in which all aspects of the interaction are kept the same across conditions except for the participant's beliefs about the nature of ~~their~~ his interlocutor [3].

Through this paradigm, we ~~could~~ can manipulate the participant's beliefs about ~~their~~ his partners and lead ~~them~~ him to think that he~~hey~~ ~~are~~ is interacting with a human or a computer while keeping other factors constant. Accordingly, any changes in the participants' behaviour ~~could~~ can be associated with the difference in their beliefs about their interlocutors. Further~~more~~, this paradigm allow us to study the interaction with the simulation of more developed technologies that not currently available yet rather than interacting with a final system[11].

Comment [Editor14]: Please check if it shoul be participant's or participants'..As a group or an individual for the sentence will change grammatically.

Comment [Editor15]: The grammar will be based on whether it is an individual particpnt or a collective gathering. Please check and specify

On the other hand, ~~on~~ in order to observe syntactic alignment proposed in this project, we need -an intelligent system -that offers ~~a full~~ a tremendous degree of ~~sensitivities~~ sensitivity as required to achieve this observation-; hence such system can~~is~~ not exist yet, nor ~~using~~ use an automated system (rather using a controlled one) without risking the experiment to committing possible errors when the computer takes ~~their~~ its turn, using the Wizard~~a wizard~~ of ~~oz~~ Oz method, which would be the best alternative.

2.1 Experiment Design

Taking into account the experiment design used in [11] and ~~also~~ the use of the (PO) and the (DO) structure used in [7], [10], [13], the same collaborative communication game was employed in this project in which two players (the naïve participant and a confederate) ~~take~~ took turns in describing and matching different picture cards using a picture matching game application. In keeping with the three different conditions used in [11], the confederate (either basic voice computer, advance voice computer or human) produced mixed scripted utterances of the (PO) and the (DO) structures and the target utterance produced by the participant ~~afterword~~ afterwards was observed to detect whether alignment ~~has~~ had occurred or not.

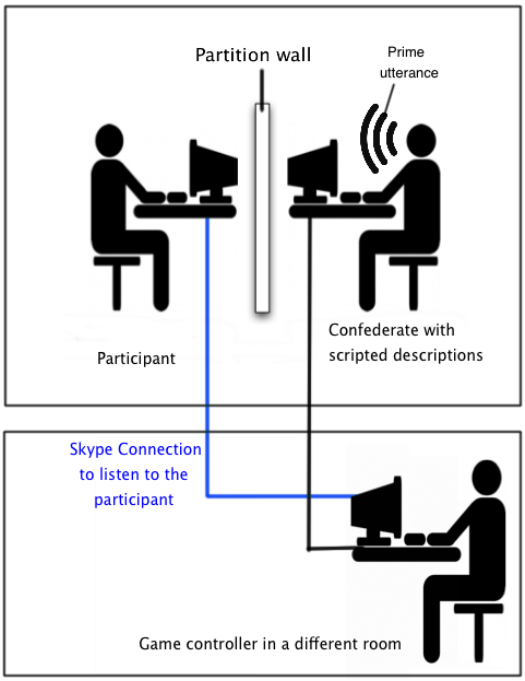


Figure 1. Experiment Set Up for the Human Condition

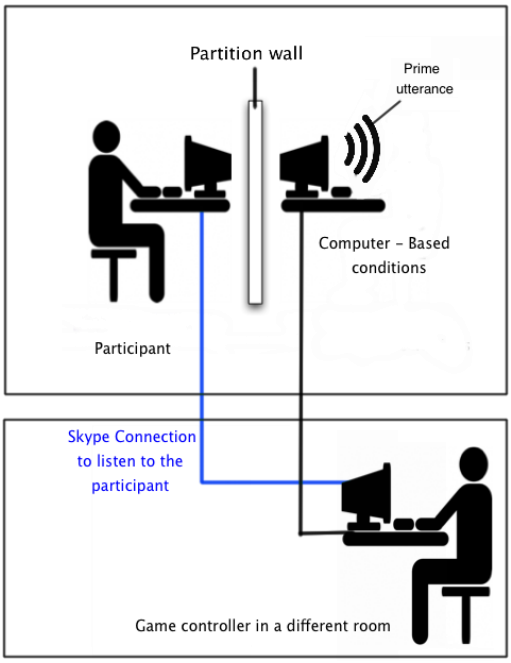


Figure 2. Experiment Set Up for the Computer-Based Conditions

2.1.1 Conditions

The experiment ~~contains~~ contained three different conditions in terms of the naturalness of the partner's voice in which they acted as ~~an~~ independent variables to compare alignment occurrences between them. The human condition (first condition) was used as a control to compare the effects between human-human utterances ~~to~~ with computer-human utterance. While the computer-based conditions~~,~~, the basic computer voice that used in the game as *basic* condition (second condition) was an artificial sounding computer voice and it distinguished by the absence of naturalness in intonation and more of ~~and~~ automated generated speech. The descriptions for this condition were produced using the Vox Machina for Mac. However, in the *advance* computer voice condition (third condition)~~, the~~ participants ~~were interacting~~interacted with a computer that ~~produces~~ produced more ~~human~~ human-like and natural utterances. To produce these more natural utterances, another program ~~were~~ was used called the Festival Speech.

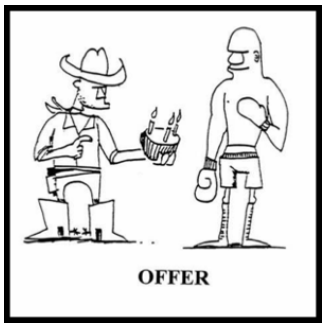
Comment [Editor16]: The participant? Naive confederate?

Comment [Editor17]: Please check for clarity

Comment [Editor18]: What is the full form fo this? Is it an operating system?

2.1.2 Materials

Each primed picture in the game and its description followed by a second picture was to be described by the participant and these were termed~~called an~~ items. There were 72 items in total, 24 of them, ~~were~~ experimental ~~items~~ and derived from items originally used in [7], [12]. The experimental pictures represented ditransitives involving an agent, a patient~~,~~ and a beneficiary. Six ditransitive verbs were used (give, hand, offer, sell, show and throw). Each verb ~~were~~ was used four times as a prime picture and four times as a target picture and each prime picture with its description ~~will~~ would be followed by a second picture (target picture) to form one experimental item. The remaining 48 items were filler items whereby each filler item consisted of a prime and a target picture but only ~~representing~~ represented the transitive actions involving an agent and a patient (e.g.~~,~~ a waitress eating a cake). In the filler pictures, 18 different transitive verbs were used, and each verb was used at least twice as a prime and ~~at least~~ twice as a target. An example of an experiment picture and a filler picture ~~was~~ is shown in Figure 3. We constructed two lists containing 24 experimental items and 48 filler items. Each list contained 12 experimental items with PO target descriptions and ~~with~~ 12 with DO target descriptions. The two lists were similar in terms of the order of the items, but the only difference was the experimental items that were used with a PO description in one list and ~~were used~~ with a DO description in the other. The same ~~and the same were~~was applied ~~on~~ to the experimental items that were used with a DO description.



Experiment picture with a ditransitive verb

Filler picture with a tansitive verb

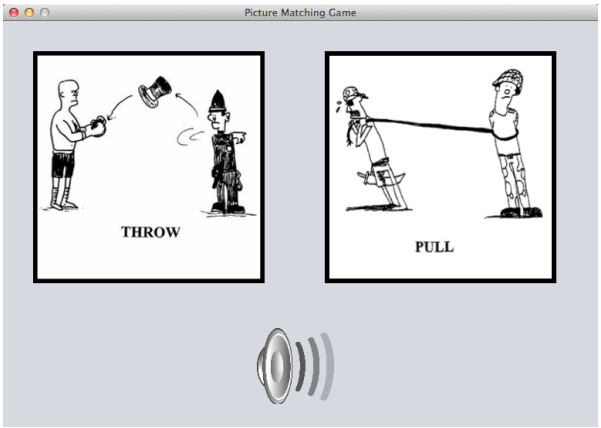
Figure 3. An example of the experiment and the filler pictures.

2.1.3 Controls

In the experiment, several controls were used to ensure that any priming effect was ~~n't~~ not attributed to ~~due to~~ other factors. To eliminate any priming effect caused by lexical similarity between primes and target descriptions, the verbs used in the target picture ~~is~~ were always different from the ones used in the target. Also, adding ~~2~~ two filler items after each experiment item would mean controlling the potential of ~~carry~~ carrying-over the priming effects and ~~to~~ hideing the focus of the experiment item descriptions. Additionally, in the matching turns for the participant, the prime picture was paired with a random filler picture or a picture with a ditrasitive verb different from the ones that were used in the experiment items besides a random appearance of the prime picture to be on the left side or on the right side of the game frame. Finally, a random order ~~were~~ was generated for the 24 experimental items and each one was followed by ~~2~~ two filler items. Then ~~then~~, this order was ~~hard~~ hard-coded

Comment [Editor19]: What could these be?

Comment [Editor20]: Ditransitive?



across the experiment's ~~3~~ three conditions to certify that each participant ~~went~~ had gone through the same settings-.

Figure 4. Example screenshot of the communication game where participant plays the role of the matcher.

Figure 5. Example screenshot of the communication game when the participant turn to describer.

2.1.4 Participants

~~44~~ Forty-four participants including 24 females and 20 males, all ~~were~~ members of the University of Birmingham community were paid to participate in the experiment.

2.1.5 Procedures

Participants were randomly assigned to one of the three conditions taking into consideration equal distribution for males or females between them. They were told that ~~It was explained to them that~~ they ~~are~~ would be ~~going to~~ playing a picture matching game and they ~~will~~ would have to pick out the picture that matched a description they ~~will~~ would hear about from their partner and ~~as~~ quickly and as accurately as possible ~~and~~ they would take turns to~~then take a turn to~~ describe one to their partner. In the *human-interlocutor* condition, the participants were told that they ~~were~~ would have to~~to~~ play the game with another participant (confederate) on the other side of the room through a network computer terminal. ~~While i~~In the *computer-interlocutor* condition, participants were shown the computer on the other side of the room and were told they are going to play the same game with a computer placed on the other side of the room via a network terminal and that it could 'understand' descriptions of pictures

Comment [Editor21]: The picture?

~~placed on the other side of the room via a network terminal~~. All other details of the experiment procedure were identical. Before starting the game, the participants were asked to complete a demographic form recording their age, gender, profession/subject for study and ~~they~~ were asked if they were ~~a~~ non-native speakers of English ~~speaker~~ or if they suffered from any medical condition that affected their ability to safely view the computer screen. If one of their answers was 'Yes' ~~to either of the two questions~~ they were told that they ~~are~~ could not participate ~~not qualified to take part~~ in the experiment ~~and~~ though they were thanked for showing interest in the study. Participants underwent ~~through~~ a trial run consisting of ~~4~~ four items so they could get familiar with the game. During the game, the experimenter noted participants' utterance structures in addition to an audio recording ~~to~~ of the whole session to guarantee that we could retrieve it if the experimenter missed it during the session. Each experiment's session ~~were~~ was controlled remotely by a member of the experiment team using Skype to listen to the session and a remote desktop control session was offered ~~from~~ by subscribing at Logmein Rescue website to control the laptop in the experiment room to ensure that the game turns for participants and their partners were correctly taking place by clicking the ~~*next*~~ *next* button after the participant ~~have~~ had picked up a picture when it ~~is~~ was the participant's turn as a matcher or after the participant finished his or her description~~from description when he/she turn as a describer~~. The remote desktop control session was also used to play the audio file that joint to the prime picture when the confederate (human or computer) needed to play the role of a describer. After completing all the 72 items, they were thanked and debriefed about the nature of the experiment.

Formatted: Font: Italic

Comment [Editor22]: Please check for clarity