University of Potsdam
Seminar “Interaktionsdesign für soziale Roboter”
Docent: Maike Paetzel-Prüsmann, Ph.D.
Summer Semester 2020/21

Pentomino with Proto

Group 3
Wencke Liermann
Lisa Plagemann
Niklas Stepczynski

Contents

1 Introduction

1.1 Pentomino with *Proto*

The following report describes the process of developing a human-robot-interaction (HRI) using a model of the anthropomorphic *Furhat* robot by *Furhat* Robotics AB which we named *Proto*. We used the *Furhat* software development kit (SDK) and a base implementation of the Pentomino game to implement our interaction model. Our goal was to enable a naive user to build a 12-piece elephant figure on a virtual playing board out of Pentomino pieces using only voice commands in a dialogue with *Proto*.[1]

1.2 Dialogue Model

For the task of having *Proto* play a game of Pentomino with the user, we decided to embed the game in a casual dialogue with a greeting and finalisation phase. The obvious goal was to enable the user to complete the game successfully while creating a friendly, polite conversational atmosphere. When creating *Proto's* voice lines, we were mindful to communicate a sense of cooperation and guidance through the dialogue while keeping the conversation light-hearted. We used voice lines to communicate what kind of input is expected from the user and to give encouraging feedback in the course of the conversation.

While not interacting with any user, *Proto* will go into an idle state in which randomised gestures make the robot seem animate and approachable. Whenever a new user enters *Proto's* field of view, the robot will offer a greeting. We managed the option of several users in close proximity by only concentrating on one user at a time - the one that entered first - in order to keep the dialogue manageable. After the user signals the willingness to communicate by responding when greeted, *Proto* will suggest to play a game. We decided to keep the amount of small talk to a minimum since the primary goal of our conversation is the game itself. Therefore we kept the introductory part of the conversation rather basic so that it could be handled with simple yes/no-responses of the user. Once the user agrees to play the game, *Proto* will give an explanation of rules and options for the game. The user then has to confirm that the explanation was understood or ask for it to be repeated. An inexperienced, new user will always be sent into a trial mode of the game first. They may only progress to the real game once this trial was successfully completed. While the game is running, every game round will start with *Proto's* request for information on a target Pentomino piece. If the user fails to give sufficient information for an unambiguous selection, *Proto* will verify the integrity of the shared information before asking for additional input. Once enough information for an unambiguous selection is gathered, *Proto* will suggest said piece to the user by highlighting it. The user has the option to reject the offered piece which will reset the shared knowledge and remove the highlighting. *Proto* will also suggest and highlight a random piece if the user does not provide any information for too long. Upon accepting the highlighted piece, *Proto* will place it on the right board. The user will then give directions on how to navigate the piece towards its intended position. If the user hesitates to give directions at this point, *Proto* will offer a hint by highlighting the corresponding shape in the template. Once the piece is placed in its goal position, *Proto* will return to gathering information in the same way until all pieces are placed or the time runs out. In both cases, *Proto* will wrap up the game and suggest a new one which the user can accept. If the user denies the request for a new game, *Proto* will say goodbye, return to the idle state and await a

[1] Our implementation can be found on GitHub
https://github.com/wencke-lm/PentominoWithProto

new interaction. A new user will have to go through the whole entry phase (including greeting, explanation and trial) while a known user will be directly invited to play a game.

During the dialogue *Proto* will establish eye contact when talking to the user or deliberately look at the object of interest, such as the board with the selected piece, to support the dialogue flow and turn taking. For a comprehensive state graph including *Proto's* gaze behaviour, see Figure A.1.

1.3 Measuring Dialogue Quality

For testing, we decided to define a number of criteria to measure the dialogue quality of our implementation. In order to test our criteria, we presented 8 naive users with our implementation and asked them to play one Pentomino game with *Proto* while we acted as an observer. Our test group consisted of 1 female, 5 male and 2 gender diverse participants aged between 17 and 33 with English proficiency ranging from a B1 to a C2 level.

Due to social distancing restrictions of the current situation, we decided to compromise on testing with the *Furhat* SDK at each of our private set-ups. For an example of the experiment set-up, please refer to Figure B.1.

Since the Pentomino game can be seen as a cooperative, goal-oriented process, we wanted to measure the attainment of said goal on the one hand, and the perceived dialogue quality by the user on the other hand. In order to enforce equal conditions for all participants, the only instructions provided priorly to the actual interaction with *Proto* were those included in the *Letter of Acceptance & Declaration of Data Protection*. There, we also asked for the user's consent to audio recordings of the conversation, the recording of the screen and the experiment itself. For the measurement of dialogue quality, we consulted the Godspeed questionnaire (Bartneck et al., 2009). We selected the attributes we found fitting for our implementation and presented them to our participants in the form of the *Questionnaire for the User*. Since we did not test with the real robot, we disregarded the Anthropomorphism category. Ratings for aspects like "Moving rigidly" vs. "Moving elegantly" would have been neither meaningful nor informative. We also did not include the category of *Perceived Safety* in our set-up as we could not think of a reason why the user should feel threatened by the SDK. Additionally, we defined some implementation specific criteria listed on the *Questionnaire for the Supervisor*, those were also meant to cover aspects of the feasibility assessment. All documents used for testing can be found in Appendix C.

2 Implementation process

2.1 General Modifications

We decided to change certain default settings. In particular, we customised the default texture *Marty*. The *Marty* texture was modified through the addition of clearly visible eyebrows in order to amplify the expressiveness of implemented gestures. Also, the eyes of the texture were substituted with a clearer version including more visible light reflexes. The aim of this modification was in establishing incontrovertible eye contact with the user.

The voice of *Proto* was set to *Matthew-Neural (en-US) - Amazon Polly* which is a neural text-to-speech (NTTS) voice provided by Amazon Polly. The NTTS voices provide higher speech quality (Amazon Web Services, 2021) as well as expressiveness and naturalness (Simon, 2019). The *Matthew-Neural (en-US) - Amazon Polly* voice was selected due to its perceived

friendliness in the conversational style with the aim of aiding a natural dialogue atmosphere.

The default idle behaviour of *Proto* includes blinking but no movement of the head itself. In order to make the robot more lively, we assigned randomised gestures to the idle state. While *Proto* is not attending any user, the robot will look around the room, let his eyes gaze away or look down for a moment. Our goal was to make *Proto* look more approachable since we felt the initiation of a conversation of an unmoved robot might feel strange to the user. Additionally, we used default gestures as well as custom gestures in the dialogue itself to make *Proto* appear more responsive and to aid turn taking. For example, when *Proto* fails to gather sufficient information, the robot will display confusion to accompany the assigned voice lines. We also made sure to show fitting reactions to the specific development of the game, such as the user winning or loosing the game, in hopes of creating more lively reactions from *Proto*.

We decided against the implementation of any form of name recognition. While using the user's name in the dialogue might communicate intimacy or individuality, the implementation of name recognition proved to be failure-prone. Due to the sheer number of possible names and difficulties in processing non-English names, we decided not to ask for the user's name. Nevertheless, *Proto* will recognise the user by the ID assigned in the SDK. We used this mechanism to handle a possible interruption of the dialogue by any additional user. If *Proto* recognises a new user, while being engaged in a conversation, *Proto* will kindly ask them, not to disturb the conversation. While humans can manage a change of conversation partner naturally (Sacks et al., 1978), we wanted to ensure that *Proto* can focus on one person. Dealing with the third party in this way might do both, keep distractions for the voice recognition to a minimum and showcase a way in which *Proto* can react to the surroundings. Additionally, we used the user ID to manage custom behaviour towards different users. Did a user once decline *Proto's* invitation to play a game, they will not be asked again as long as other users are present. If a known user remains present in *Proto's* field of view, they will be asked whether they had a change of mind, showing that *Proto* remembers them.

2.2 Selection Process

2.2.1 Explanation and Trial Mode

In our implementation, *Proto* will initiate the conversation with a user who enters *Proto's* field of view. After greeting the user, *Proto* will enquire whether the user wants to play a game. If the user agrees, *Proto* will explain the game to them with the following instruction:

> *"This Game is called Pentomino. The goal of the game is to build a shape out of the pieces you see on the left. Please describe the pieces one at the time. You can refer to the piece's color, position on the grid or by comparing the shape to a letter. We need to move each piece to the correct position in the shape. To move a piece, specify the direction and tell me when to stop. You might have to rotate or mirror the pieces to make them fit."*

If the user states that he understood, *Proto* will guide the user to a trial run of the Pentomino game. We decided to add a simple trial version of the game to present the user with an easy introduction to the

mechanics. In the trial version, the user is asked to build a simple square on a template out of the P, U and V Pentomino pieces. We chose those pieces to encourage the user to describe their rather approachable shapes. Since we expected the selection process to be easier to understand, we wanted to focus on giving the user the chance to experiment with the placement process. We set the game timer to 300 seconds for the trial mode so that the user had time to test the mechanics sufficiently. Only if the user has finished the trial mode successfully, we allow them to progress to building the elephant. We decided to implement this requirement to ensure that the user has had enough practice in talking to the robot. Failing the trial mode will trigger *Proto* to give further hints to the user, alluding to the most common issues in the following way:

> *"You know - I might have some difficulties understanding. Please, make sure to speak loudly and clearly! If you use short but complete sentences, I might be able to understand you better. I will try my best to follow your instructions. You are doing great!"*

Since we were aware of certain shortcomings, we wanted to communicate some further advice for speaking with the robot. Also, by making *Proto* acknowledge these issues, we wanted to ensure that the user will not feel discouraged.

Proto is supposed to explain everything about the game autonomously, so that a naive user could be guided only by the robot itself. When presenting the implementation to naive users, we purposely gave no explanation about the game so that we could test whether *Proto's* explanation and the trial run were sufficiently instructive.

2.2.2 Selection Criteria

In the first part of the Pentomino game, the user is asked to select a piece from the left board. Since the selection should follow the user's voice command, a number of different phrasings needed to be taken into consideration. In *"wizard of oz"*-style trials, we identified the attributes shape, colour and position on the board, as well as the position of a piece in relation to another piece, as common topics in the users' descriptions. In order to filter out the relevant information in the user input, we implemented custom intents for said attributes which were accessible through the system's natural language understanding (NLU) component.

For the English language, earlier research found evidence for a preference of shape in descriptions of referents of varying materials (Imai & Gentner, 1997). Since the Pentomino pieces resemble simple shapes with no apparent function, a comparison to suchlike findings might be a basis for establishing the hierarchy of user contributions. Given that the experiment is aimed at a natural English conversation, we decided to accommodate the language-specific pattern of the "habitual attention to shape" (Imai & Gentner, 1997, p. 190). Imai & Gentner (1997) also found a focus on plausible names or functions in the description of simple shape objects by American adults. Possible references to a distinct function or object that is comparable to a Pentomino piece were accounted for in the NLU SHAPE intent. We also included the input of further functional descriptions from fellow students (Kröner et al., 2021) working on the same project. In the specific case of the Pentomino game, the shape of the pieces holds an important status since the correct identification of a SHAPE intent will lead to an unambiguous selection. While colours and location both hold a certain range of vari-

ance in user perception, the shape of each piece is unique. The possible risk of the user misidentifying a shape due to mirroring or rotation is averted by several alternative descriptions in the NLU SHAPE intent. However, in testing, the distorted orientation of the Pentomino pieces led to errors for certain pieces. For example the N-shape and the Z-shape were interchangeable when rotated. This problem is accounted for by offering an option to ignore a part of the user's input during the selection process. An alternative choice is presented based on parts of the input that are in alignment with Pentomino pieces currently present on the board. The suggestion of a close alternative might aid both the feeling of naturalness and the overall progress of the selection process.

As a next possible attribute for the selection process, we decided to use the colour of the Pentomino pieces. The original implementation of the Pentomino game provided 11 different colours (yellow, orange, beige, light red, pink, purple, light green, dark green, light blue, dark blue and turquoise) which were randomly assigned to the pieces. The retention of those given colours created some uncertainties in the selection process. Both the user's individual interpretation and unique compilation of randomly assigned colours on a given board, held challenges for a correct identification. In order to accommodate the user's individual interpretation of given colours, we mapped similar concrete colours to a group of superordinate colours (blue, green, orange, red, yellow, beige). In combination with accounting for several colour descriptions in the color intent, this approach offers more latitude in processing the user's colour input. Based on the cooperative principles of the Gricean maxims (Dale & Reiter, 1995), we assumed that the user will specify the information of colour (and position) of the Pentomino piece in a way that is only as detailed as required by the situation. In the specific case of randomisation of the Pentomino game, a user may be represented with only one light blue piece and no other piece of the superordinate colour group "blue". In this situation, the user might want to refer to the piece's colour as blue and refrain from further specification. The implementation of superordinate colour groups therefore accounts for imprecision of user input by establishing a reasonable room for interpretation.

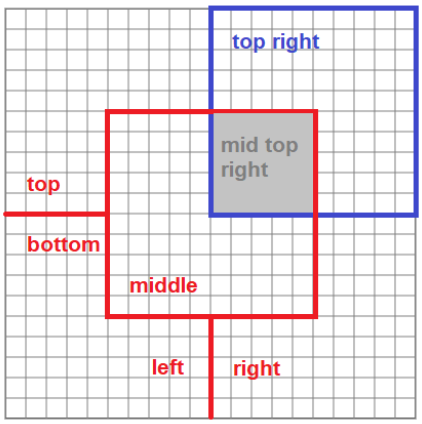

Figure 1: Board Partition for LOCATION Intent

In order to process location information in the user's expression, the Pentomino board was divided into different areas as visualised in Figure 1. Using the coordinates of evenly partitioned borders, we mapped areas of the Pentomino board to certain instances of the LOCATION intent. As implemented in the original game, the location assigned to a specific piece is achieved by referring to the upper left block of the piece rather than the whole piece itself. Due to the randomisation process, the program might therefore have placed a piece in a way where the absolute position perceived by the user does not match the position assigned by the program. In order to deal with this inaccuracy, we also gave the option to ignore certain details in the location description. This allowed us to find a suitable alterna-

tive based on more general location details in the input to offer to the user.

In an experiment by Golomb et al. (2014), participants were presented with stimuli comparable to the Pentomino pieces and had to perform a differentiation task. They were asked to base their judgement only on the task-relevant categories colour and shape while ignoring irrelevant dimensions (i.e. location). They used subtle differences for the task-relevant property and large differences for irrelevant object properties. The authors note that participants were unable to suppress the influence of location despite it being irrelevant or detrimental to the task. This phenomenon did also occur in a similar experiment by Tsal & Lavie (1988) where participants were instructed to report one out of two targets based on the colour of a cue, while asked to ignore its location. Golomb et al. (2014) did not observe anything similar for neither shape nor colour. Given this observation, it can be assumed that a human user would be more willing to ignore shape or colour information that prevents them from uniquely identifying a piece, if it in turn enables them to find a piece that matches the described location. However, a similar experiment by the same authors (Golomb et al., 2014) alludes to the fact, that this superior role of location can only persist if the differences in other object properties are not too obvious. Therefore we contribute an even more important role to the abstract colour category and assume that human users would try to meet this criterion first.

We disregarded expressions using relations of the desired selection to other pieces, such as descriptions including prepositions, in the implementation process. The identification of a secondary Pentomino piece did not seem to yield a positive trade-off after other attributes proved to be sufficient for the identification process. However, the inclusion of comparatives in the LOCATION intent was able to catch certain expressions, such as “lower” or “higher”.

In order to be able to cope with possible inaccuracies or ambiguous situations, we have decided to arrange the various properties that a piece can have in a descending hierarchy of abstract colour group, location, shape and concrete colour. Based on this hierarchy, attributes identified in the user input were handled according to their ranking. Meaning that attributes which ranked lower were ignored first if no match could be found that satisfies all attributes. We assumed that the user subconsciously applies the Gricean Maxims for the benefit of the course of the game. The user will not communicate incorrect information (*maxim of quality*) and does not include unnecessary information (*maxim of quantity*) (Dale & Reiter, 1995) in the description of a piece.

While the colour makes for an obvious attribute for the user to refer to, a certain margin of error has to be accounted for as described earlier. In order to deal with this issue, the abstract colour is considered first in the hierarchy. Based on the Gricean Maxims, a specification of the concrete colour should be taken into account when necessary. Placing the concrete colour at the bottom of the hierarchy, this information can be processed when required.

Both the significant value of the location of an object described in research and the accessibility of a location description in the user input were our motivation of including the LOCATION intent second in the hierarchy. While there is evidence for the shape description of simple objects to play an important role in identification, the specific shapes presented in the Pentomino game where of varying distinctness. The SHAPE intent takes the third place in our hierarchy in order to process shape information when useful.

2.3 Placement Process

2.3.1 Realising Movement

Similar to the selection process, we gathered examples of how to describe the placement of a Pentomino piece on its designated place in the elephant shape. First of all, we had to account for options to rotate and mirror the Pentomino piece. While we gave the option to specify the direction and the exact degree of the rotation, we used a rotation of 90° clockwise as the default value when a simple ROTATION intent was identified. The value for degrees was mapped to multiples of 90 to aid the game process. For the MIRROR intent we assigned the default behaviour of mirroring the Pentomino piece on the vertical axis.

For describing the movement of the Pentomino piece on the board, we found several different varieties of possible user input. One option was to give a direction and wait until the Pentomino piece has reached the desired position and cancelling the movement with a stop signal or a new direction. Another option was a precise information about the desired distance and direction of the movement. We decided to account for both options in our implementation. To realise the first option, we employed incremental speech recognition. *Proto* would process an utterance as soon as received, but would continue to listen until a larger pause, a different instruction or a STOP intent was detected. Accepting the second option also gave the opportunity to make small adjustments to an already moved piece or to give precise instructions for shorter distances. Given the case that the user fails to stop the movement precisely and as a result the piece ends up one block too far to the right, the user can then ask to move it one block to the left.

The anticipated user input for mirroring, rotating and directing the movement on the board was also mentioned in the explanation at the beginning of the game (see 2.2.1).

We also observed that an option to undo the previous action would be useful. To achieve this, we stored the user's previous action and retrieved it when the NLU BACK intent was identified in the user input. We made *Proto* undo said action to undo a mistake or to cope with a delayed reaction. A corresponding command to repeat an action was caught in the AGAIN intent that we defined. Especially for a rotation it seemed more natural for the user to ask to repeat an action instead of giving the same instruction again. In the process of the game the user will therefore be able to undo or repeat any rotation, mirror or direction instruction.

In the course of testing we found it necessary to establish borders of the right board which would stop the Pentomino piece from moving outside of the board. We did this to avoid any confusion and because there was no obvious reason for having the Pentomino piece leave the board.

We decided against the use of specific location descriptions, such as "trunk" or "front leg", since only a few of those options were unambiguous due to pieces overlapping into different areas. Also, the absolute position of a Pentomino piece being defined by only the x- and y-coordinates of its upper left block, would have made the assignment to such areas rather vague. We expected simple directions to have a similar effect while also encouraging a more dynamic nature of the game.

In order to keep the game play overseeable and clearly structured, we decided to keep a strict order of selection and subsequent placement of the selected piece without the option to proceed with selecting another piece before the selected piece was placed. In this way we avoided stacking pieces on top of each other or covering up parts of the elephant shape. Given the case

that the user finds it difficult to locate the goal position of the selected piece, we considered the implementation of a hint function as described hereafter.

2.3.2 Assistance Options

For the placement part of the Pentomino game, we found it necessary to provide the user with further guidance on how to build the desired elephant shape. We decided to permanently display the elephant shape built out of copies of the playable Pentomino pieces on the right board. The copies are arranged in the desired pattern and uniformly grey coloured. While the borders of each Pentomino piece are visible, a naive user might still be confused on where to find the correct position and orientation for the piece which was selected. If the user does not provide any instruction for too long, *Proto* will offer help by highlighting the copy in the template which matches the selected piece. Additionally, the user can express a need for help. We anticipate that this hint function acts as a form of deictic gesture which has been identified as being especially useful in identifying objects unambiguously and fast (Sauppé & Mutlu, 2014). Our aim was to support the course of the game within the given time limit and to prevent the user from becoming confused or even frustrated.

Since the original implementation of the Pentomino game did not offer a template, we decided to keep it as an option. When building the game, the template can be deactivated. We decided to keep the hint function since an experienced user might still want some orientation on the board.

3 Discussion

Certain user expressions including relations of a described Pentomino piece to a secondary piece, such as phrases using certain prepositions or comparatives, were not accounted for by our implementation. In order to catch that type of expression, the LOCATION intent could be extended. It might be necessary to exercise caution not to create unwanted interruptions in the dialogue due to processing time. Also, expressions in the style of *"move the piece to the right by 3 blocks"* resulted in a movement that had to be corrected once the user realised that the movement did not stop at the desired location. This was caused by the use of incremental speech recognition. We accepted this shortcoming in favour of a faster, more intuitive movement since the mistake could be easily corrected.

The predefined colours of the default game configuration posed a problem as it included several colours of the same abstract type (e.g. turquoise, light blue) as well as colours diverging noticeably from prototypical representatives (e.g. light red). The problem was addressed by the definition of groups containing similar colour shades (e.g. blue). This approach allowed some flexibility in processing the user input and proved satisfactory in testing. After presenting the game implementation to different users, it became apparent that the room for interpretation for colours is rather wide. In several cases the randomisation of colours left the user with a game board that led to generous generalisations: with no other blue piece on the board, a turquoise piece was described as blue or with no orange piece present, the user described a beige piece as orange. While our implementation was able to catch the first description, the second case was unsuccessful and led to some irritation for the user.

It should also be taken into account that colours might be displayed differently on different monitors. Those inconveniences could be easily avoided by defining colours for the Pentomino pieces which are easily distin-

guishable and which are not prone to colour distortion on varying systems. Additionally, colours prone to problems with identification (colour deficiency or language-specific colour terms (Winawer et al., 2007)) should be avoided. For a possible colour modification of the game, focal points of colour groups (Kay & Maffi, 2013) could be considered in order to define a fitting representative for a colour.

The current state of the NTTS voice *Matthew-Neural (en-US) - Amazon Polly*, which we used in the implementation, lacks certain modification possibilities. We found the lack of options for a happy or sad voice style limiting in modelling emotional nuances in the dialogue. While we perceived the NTTS voice as friendly, we would have liked to adjust the pitch.

The implementation of gestures seemed to be a valuable addition to the dialogue flow. The default gestures covered basic emotional states but we found them to be rather unnatural. While trying to implement custom gestures to convey the emotions we were aiming for, we found the customisation process to be difficult and time-consuming. The implementation of natural gestures via the provided parameters proved to be challenging and we discussed the need of a broader library of gestures or the option to create gestures with puppeteering.

It would have been a nice addition to ask the user for the desired level of difficulty, i.e. whether they want to disable the template. However, this would have required an additional event on *Proto's* side and we decided to limit alterations to the communication channel to a minimum.

In testing, we found that users had their own interpretation of the rules which *Proto* presented to them. At the beginning of the demo mode which we implemented, some users gave detailed descriptions of the whole placement process or even the selection and placement process in one large input statement. Since the robot only processed a small chunk of the information, the user seemed to adjust to this behaviour rather quickly and modified their input accordingly. This was proof to us that the implementation of the trial mode was a useful addition to the game since the user was able to practise and adjust their instructions. This also helped to save time in the game itself by equipping the user with the necessary means of both talking to *Proto* and playing the Pentomino game. Enriching the explanation by a *"learning by doing"*-part therefore seemed to be a useful addition.

Despite offering a large number of expressions in the definition of our custom NLU intents, we did not manage to account for all the expressions that where used by our participants. Some valuable additions to the MOVE intent could for example be the *square* and *space* part in sentences like "Move it one space/square up." This shows that even as a group of three it is challenging to create comprehensive custom intents manually. One possibility to address this issue would be to refer to domain specific corpora. For the relevant domain those, however, seem to exist only for the German language.

We decided to create our implementation for the English language. Despite both our test group and ourselves being native German speakers, we made this choice due to the advantages of the English syntax and morphology. Both in the process of implementation and in its testing, it became obvious that the automatic speech recognition (ASR) did not always capture the user input correctly. We came to realise that the processing of the input of a non-native speaker is challenging for the system. We tried to counteract common ASR mistakes by including them in the respective intent (e.g. *peace* for *piece*). Nevertheless, all of our

participants had to deal with *Proto* misunderstanding them. Some of these misunderstandings led to mistakes in the game. As displayed in Figure 2, out of an average of 8.75 mistakes in the selection phase 5.875 mistakes were accounted for by the ASR. For the placement process, ASR accounts for 6.5 out of 6.875 mistakes. We argue that the higher number of mistakes in the selection process is due to a great variability in actually presented descriptions. Whereas during the placement process participants simply followed their established pattern of instructing *Proto*.

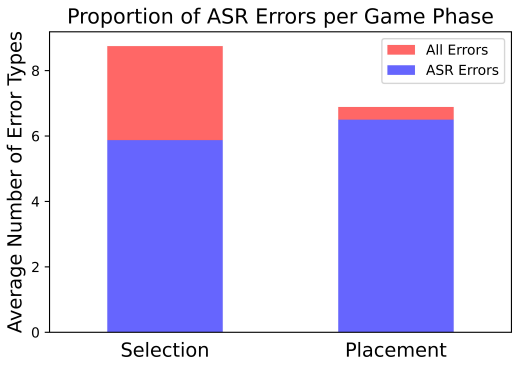

Figure 2: Proportions of ASR Mistakes

In testing, we found that the feedback we used to structure the dialogue came at the expense of time. Despite keeping the voice lines short, certain responses (misunderstanding, clarifying information) took a considerable amount of time. In combination with ASR mistakes, this led to a disruption in the game process, causing 6 of 8 participants to run out of time. Yet, our participants placed 9.5 pieces on average and 2 participants won the game within the time limit, proving that it is possible to win the game with our implementation.

The feedback from our participants in the adapted Godspeed questionnaire (Figure 3) shows positive average ratings in the categories *Likeability* and *Perceived Intelligence*, while the greatest potential for improvement was seen in the *Animacy* category.

We were pleased with overall positive ratings in the *Likeability* category. Since we tried to modify *Proto* to create a pleasant conversational partner equipped with friendly, polite voice lines and gestures we found our efforts validated. We managed to establish a strong friendly baseline that persisted even behaviour normally considered impolite. Despite the users and *Proto* interrupting each other 7 times on average, Proto was not perceived as rude. Interruptions became more common towards the end of the game especially when the user started to anticipate and dismiss *Proto's* repetitive reactions, such as when asking for a next description or in cases of misunderstandings. This might point to an effect of becoming accustomed with the game and voice lines or simply to the user running out of time. Both possibilities could be looked into in a revision attempt.

Our participants described *Proto* as leaning towards machinelike and artificial. In our implementation we focused more on the goal of winning the game. It might be a valuable addition to incorporate more non-goal-orientated conversation aspects. For example, *Proto* could engage in more small talk before initiating the game. We also assume the *Animacy* category would be influenced towards a more positive rating if the participants were presented with the real robot instead of the SDK. Luria et al. (2017) found that the embodiment of a robot supports a more enjoyable conversational atmosphere. We argue that our HRI would also benefit from this level of proximity.

Due to the current situation, we decided to refrain from testing a larger number of people at one designated testing spot. We compromised on testing the implementation on each of our home systems with people in our closest circle. This compromise came

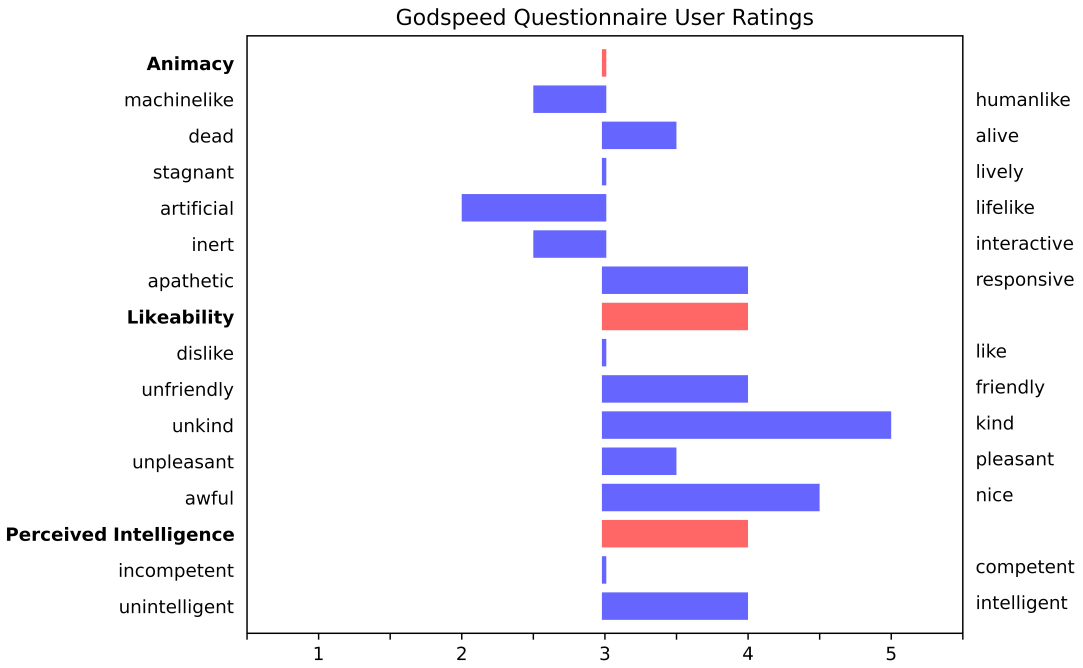

Figure 3: Adapted Godspeed (Bartneck et al., 2009) Questionnaire User Ratings

with certain irregularities in testing which we were willing to accept in order to keep any risk and inconveniences for our participants and ourselves to a minimum. We are aware that our test group lacks both size and diversity. Also, the variation of our different experimental set-ups might have established different conditions for the participants. The conclusions that we were able to draw are vulnerable to a number of biases and can only give a rough tendency.

Implementing a HRI that enables a naive user to win a game of Pentomino, strictly based on the instructions of the robot, was one of our explicit goals. We were satisfied to achieve this goal by having some participants win the game. Our implementation proved to be solid and we received a number of positive user reactions. In our implementation, *Proto* was already perceived as overall very likeable in the remote testing set-up using only the SDK. We see room for improvement towards a more time efficient dialogue flow with a few additions to our custom intents. It would be interesting to observe user ratings when presented with the robot in person to evaluate our efforts in implementing facial expressions and movements.

Overall, we have to conclude that a HRI is a complex process with a countless number of variables that have to be considered. Even for such a straightforward task like the Pentomino game, the variability coming from the user input that has to be accounted for, is hard to grasp. In order to create a creative, entertaining, natural interaction, an enormous number of factors needs to be taken into account making the implementation of a good HRI a challenging task combining various technical, social and linguistic aspects.

We believe that methods of computational linguistics can contribute valuable expertise for processing human language more efficiently and thereby improve the spoken interaction between humans and robots.

References

Amazon Web Services, I. (2021). *Amazon polly developer guide.*

Bartneck, C., Croft, E., & Kulic, D. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics*, *1*(1), 71-81. doi: 10.1007/s12369-008-0001-3

Dale, R., & Reiter, E. (1995). Computational interpretations of the gricean maxims in the generation of referring expressions. *Cognitive Science*, *19*(2), 233-263. Retrieved from `https://www.sciencedirect.com/science/article/pii/0364021395900187` doi: https://doi.org/10.1016/0364-0213(95)90018-7

Golomb, J. D., Kupitz, C. N., & Thiemann, C. T. (2014). The influence of object location on identity: A "spatial congruency bias". *Journal of Experimental Psychology: General*, *143*(6), 2262.

Imai, M., & Gentner, D. (1997). A cross-linguistic study of early word meaning: Universal ontology and linguistic influence. *Cognition*, *62*(2), 169–200.

Kay, P., & Maffi, L. (2013). Number of basic colour categories. In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online.* Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from `https://wals.info/chapter/133`

Kröner, E., Rockstroh, J., & Jenke, M. L. (2021). *gruppe2_pentominowithfurhat.* Moodle.

Luria, M., Hoffman, G., & Zuckerman, O. (2017). Comparing social robot, screen and voice interfaces for smart-home control. In *Proceedings of the 2017 chi conference on human factors in computing systems* (pp. 580–628).

Sacks, H., Schegloff, E. A., & Jefferson, G. (1978). A simplest systematics for the organization of turn taking for conversation. In *Studies in the organization of conversational interaction* (pp. 7–55). Elsevier.

Sauppé, A., & Mutlu, B. (2014). Robot deictics: How gesture and context shape referential communication. In *2014 9th acm/ieee international conference on human-robot interaction (hri)* (pp. 342–349).

Simon, J. (2019, Jul). *Amazon polly introduces neural text-to-speech and newscaster style.* Amazon Web Services, Inc. Retrieved from `https://aws.amazon.com/de/blogs/aws/amazon-polly-introduces-neural-text-to-speech-and-newscaster-style/`

Tsal, Y., & Lavie, N. (1988). Attending to color and shape: The special role of location in selective visual processing. *Perception & Psychophysics*, *44*(1), 15–21.

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the national academy of sciences*, *104*(19), 7780–7785.

A Appendix State Graph of Dialogue Structure

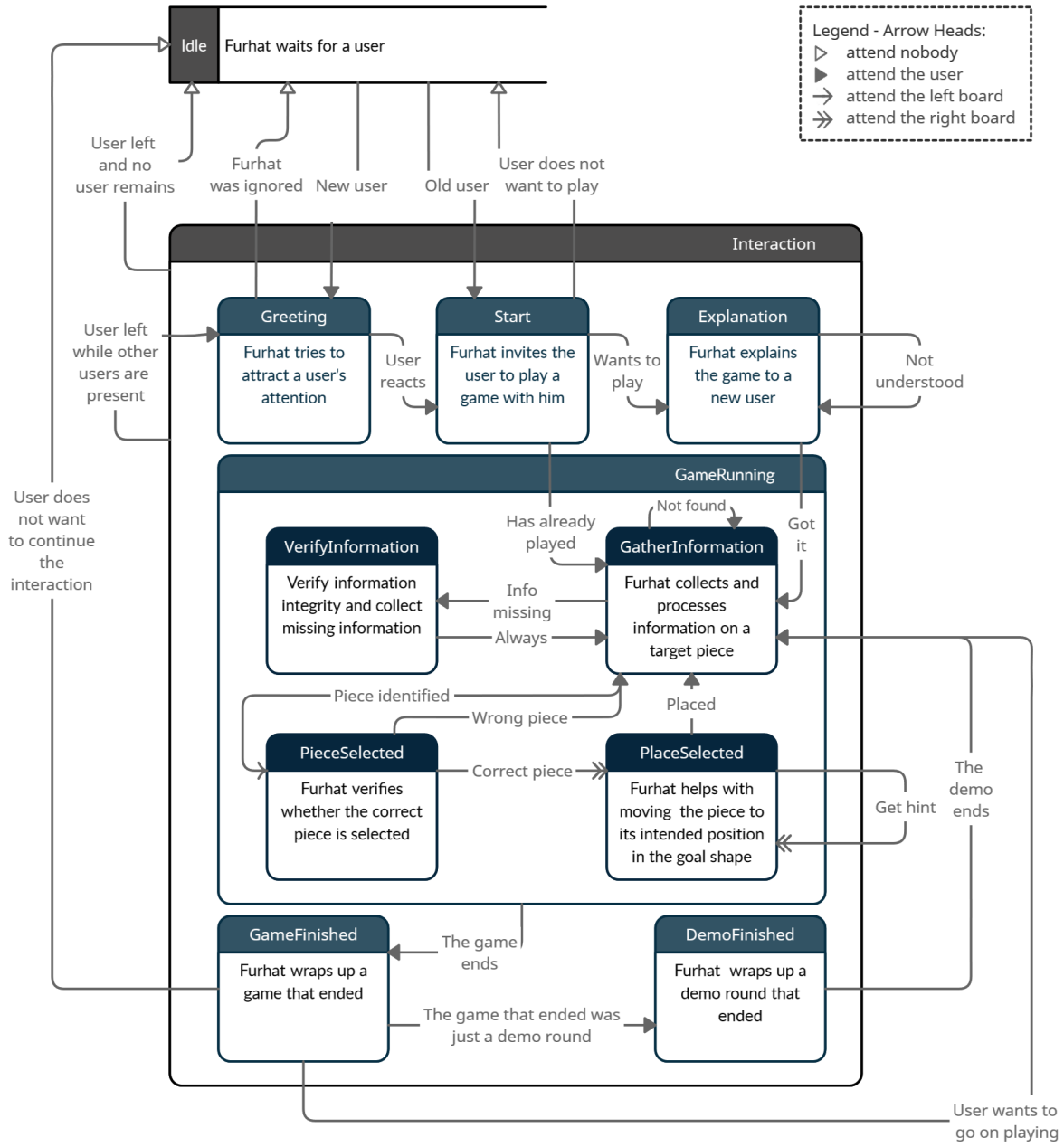

Figure A.1: State Graph of Dialogue Structure

B Experiment Set-up

Figure B.1: Example of Experiment Set-up

C Appendix Forms and Questionnaires

Please also refer to the following documents that we have attached:

- Letter of acceptance & declaration of data protection
- Pentomino with Proto: Questionnaire for the user
- Pentomino with Proto: Questionnaire for the supervisor

Additionally, scripts of our participants' conversations, corresponding images of their respective trial and Pentomino boards can be found in the folder *experiment* in our GitHub repository: https://github.com/wencke-lm/PentominoWithProto.

Letter of acceptance & declaration of data protection

to the study "Pentomino with Proto"

Thank you so much for participating in our experiment!
In the experiment, you will meet the robot *Proto*. *Proto* will invite you to play a game with him. Please agree to his offer and listen to his explanations carefully. Please make sure to speak loudly and clearly during the game! I will be present during the game but I will keep my distance and not intervene. While you are free to ask any questions you might have now, please refrain from doing so during the experiment itself.
After the game, we kindly ask you to fill in a questionnaire.

declaration of data protection:

In the process of this study, both audio and screen will be recorded. The audio recording includes the utterances of the participant as well as of the robot. While the screen is being recorded, the participants themselves will not be on camera. After the study the participant will be asked to fill in a questionnaire. The recorded data will only be processed by the supervisor and will be deleted by this person after the evaluation.

letter of acceptance:

I was fully informed about the nature, significance, conduct and scope of this study by the person responsible for the study mentioned above. I had the opportunity to ask questions and these were answered in a detailed and understandable manner before the study began. I know the form of the study and I accept its conduct.

I know that participating in this study is voluntary. In addition, I was informed that I can revoke this consent at any time and without giving reasons, without incurring any disadvantages for me.

I have been informed about the way the data will be processed and I agree that my data collected during the study will be processed in this way. I am aware that the data collected will only be used for scientific purposes. I am also aware that I have the right to have my data deleted at any time, even after the study has been carried out.

☐ I have requested a copy of this declaration of consent and have received it.

With my signature I declare my voluntary participation in this study.

_____ _____

place, date signature of the participant

_____

signature of the researcher

Pentomino with Proto

Questionnaire for the user

Age: __ Gender: __ English-Level (A1-C2): __

Please rate your impression of Proto during the game!

Submit your answer by marking a box.

	1	2	3	4	5	
unpleasant						pleasant
artificial						lifelike
dead						alive
unkind						kind
stagnant						lively
inert						interactive
unfriendly						friendly
incompetent						competent
dislike						like
machinelike						humanlike
awful						nice
unintelligent						intelligent
apathetic						responsive

If you have any questions or comments, you can write them down here:

_____

_____

_____

Pentomino with Proto

Questionnaire for the supervisor

1. Did the user win the game in under 600 seconds?

☐ yes ☐ no ☐ the game had to be aborted

2. If the game has been aborted, please state the reason

☐ Proto went into idle state unexpectedly with no prospect of returning to the ongoing conversation

☐ The user has aborted the game due to their own reasons (no interest, personal wish, incident or similar)

☐ Proto has informed the user that his speech recognition no longer works (speech recognition error)

☐ The user could not finish the demo successfully within three trials

☐ _____

3. How many pieces have been placed?

0		7	
1		8	
2		9	
3		10	
4		11	
5		12	
6			

4. How many mistakes were made in the selection process?

e. g. fn: user: "pinkies" -> Proto: "sorry, i can not find a piece"
e. g. fp: There is one piece left on the right side, user: "pick up the piece on the right" -> Proto: picks up the piece on the left side

total number: __

notes:

_____
_____
_____

4.1. How many of those mistakes were caused by speech recognition errors?

total number: __

5. How many mistakes were made in the placement process?

e. g. fn: user: "rotate it to the left" -> Proto: does nothing

e. g. fp: user: "rotate the piece to the left" -> Proto: moves the piece up

total number: __

notes:

_____
_____
_____

5.1. How many of those mistakes were caused by speech recognition errors?

total number: __

6. How often did the user and Proto interrupt each other?

total number: __

7. Has the user used expressions that are not covered by the current implementation of Proto?

_____
_____
_____

8. Other notes on the run:

_____
_____
_____