Towards Robot Learning from Spoken Language

Krishna Kodur*
kkodur@scu.edu
Santa Clara University
Santa Clara, California, USA

Manizheh Zand*
mzand@scu.edu
Santa Clara University
Santa Clara, California, USA

Maria Kyrarini
mkyrarini@scu.edu
Santa Clara University
Santa Clara, California, USA

ABSTRACT

The paper proposes a robot learning framework that empowers a robot to automatically generate a sequence of actions from unstructured spoken language. The robot learning framework was able to distinguish between instructions and unrelated conversations. Data were collected from 25 participants, who were asked to instruct the robot to perform a collaborative cooking task while being interrupted and distracted. The system was able to identify the sequence of instructed actions for a cooking task with the accuracy of 92.85 ± 3.87%.

CCS CONCEPTS

• **Computer systems organization** → **External interfaces for robotics**; **Robotic autonomy**; • **Human-centered computing** → **Accessibility technologies**.

KEYWORDS

robot learning, natural language understanding, assistive robots, collaborative robots

ACM Reference Format:
Krishna Kodur, Manizheh Zand, and Maria Kyrarini. 2023. Towards Robot Learning from Spoken Language. In *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction (HRI '23 Companion), March 13–16, 2023, Stockholm, Sweden.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3568294.3580053

1 INTRODUCTION

People with disabilities face challenges in performing Activities of Daily Living (ADLs), such as eating, dressing, getting into or out of a bed or chair, taking a bath or shower, etc. According to the Centers for Disease Control and Prevention [1], 61 million adults, one in four adults in the United States, live with a disability. According to World Health Organization (WHO), by 2050, the world's population of people aged 60 years and older will double (2.1 billion) [2].

Robotic systems have the potential to support people with disabilities and the elderly in performing ADLs. Chung et al. [3] conducted a study to investigate the most important tasks that a robotic manipulator can perform to help people with disabilities. The study found that eating and preparing meals are highly prioritized tasks. Preparing a meal or cooking themselves is a very crucial task for people with disabilities and the elderly [4]. The question arises *"How can a robot assist a person with disabilities (end-user) to prepare a meal while the end-user stays in control?"*.

Robot Learning from Human Demonstrations [5] and human-in-the-loop [6] approaches enable end-users to be part of a collaborative robotic task and remain in control. Moreover, the interaction between the end-user and the robot is important to ensure that it is not time-consuming and cognitively demanding [7]. Interacting with the robot naturally using spoken language can enable more effective human-robot interaction. For example, in a recent study, people above 65 years old preferred the use of spoken language as the most intuitive way of communication with a robot [8]. However, the use of spoken language is still limited in robotics [9].

Figure 1: User study setup for robot learning from spoken language.

In this paper, we propose a robot learning framework that enables a robot to generate actions for a task from unstructured spoken language. To demonstrate the effectiveness of the proposed work, a small study was conducted with 25 participants. The participants were asked to speak to a robot and teach it what actions the robot should do in assisting them with meal preparation, as shown in Fig.1. The robot was able to use the unstructured spoken language of the user and learn the necessary actions it is supposed to perform for preparing a meal.

The paper is divided into the following sections; Section 2 discusses the related work, Section 3 presents the robot learning framework, Section 4 discusses the preliminary results and Section 5 concludes the work.

*Both authors contributed equally to this research.

HRI '23 Companion, March 13–16, 2023, Stockholm, Sweden

ACM ISBN 978-1-4503-9970-8/23/03…$15.00
https://doi.org/10.1145/3568294.3580053

2 RELATED WORK

Robot Learning from human speech has the potential to enable non-expert users to teach the robot desired tasks in a natural and intuitive way. For example, Shao et al. [10] proposed a framework that enables a robot to perform different object manipulation tasks, such as "Put a cup in front of the bowl". The proposed robot framework consists of an architecture consisting of Bidirectional Encoder Representations from Transformers (BERT) [11] and ResNet [12] deep neural networks to achieve this. Input to the model is a natural language instruction and an image of the initial scene, and the output of the model is the robot motion trajectory to achieve the specified task. However, the proposed framework required a fixed-templated language from the Something-Something dataset [13]. A fixed-templated language is a set of predefined words and sentences that the user can use to communicate with the robot. For example, to instruct the robot to bring an object, the user must use the format "Bring me the <object>", where <object> can only take words, such as bowl or cup. The robot is not able to interpret the action that it needs to perform even for a slight modification in the instruction, such as "Give me the <object>". Shao et al. aim to learn only simple low-level pick and place actions but not high-level tasks, such as "making tea", which requires various consecutive low-level actions. To address this issue, Giorgi et al. [14] propose a method to learn high-level tasks, such as "making tea". The authors use the "You Only Look Once" method (YOLO) [15] to detect the object and use fixed-templated language to enable communication between the robot and the human. A visual representation of the object is then stored in the robot's memory. To learn low-level tasks, such as grasping an object, the state of the robot motors and the word "grasp" is stored in its memory. The robot executes a predefined trajectory whenever the user instructs the robot to perform the "grasp" action. Each high-level task is stored as an array of low-level actions; for example, to make tea, the sequence of actions that the robot would perform would be "_mug_grasp_lift_table_drop", "_bottle_grab_lift_mug_pour", "_teabag_grab_pickup_mug_throw". Similarly, Unhelkar et al. [16] proposed "CommPlan", a computational framework that decides if, what, and when to communicate with the human during human-robot collaboration. The CommPlan was used for a meal preparation task where humans and robots communicate in making a sandwich. The CommPlan predefines a fixed-templated language that humans and robots can use to communicate. However, the framework focuses on when the robot and the human can communicate in a seamless way rather than the robot learning from human instructions. Additionally, several researchers developed robot learning frameworks for industrial robotic applications, such as assembly [17] or object delivery [18]. For example, Li et al. [18] propose a robotic system that the robot is instructed by the user to perform either of these four tasks: Go to a location, deliver an object from point A to Point B, work in assembly and relocate the objects in a scene. The authors benchmark the popular language models, such as GPT-2 [19] and GPT-Neo, to find out the task recognition accuracy is 86.6%. In another study by Ahn et al. [20], a method called SayCan was proposed, in which a large language model was trained to select from a predefined set of 101 low-level tasks, such as "pick up the sponge" for execution by a robot. However, in a realistic kitchen setting, the presence of a wide variety of objects necessitates the ability to accurately identify and retrieve various objects. The limitations of a model that can only select from a limited set of 101 tasks could thus restrict the overall functionality of the robot. The authors also do not address scenarios in which a person poses unrelated questions, such as "Where is the table" and how the model would handle such instances.

The presented works in this section use fixed-templated language to be spoken, so that intent of the sentence is fixed beforehand, and the system does not need to interpret it, which is not a natural way of communication. However, cooking is an unstructured task. Persons cooking a meal at home do not necessarily follow a predefined plan for the whole process, from choosing a recipe to planning a meal [21]. On top of being an unstructured task, the system should be able to identify if multiple instructions are clubbed together. E.g., "Bring me the salt, pepper, and seasoning." In this case, the robot has to perform three pick-and-place actions each for salt, pepper, and seasoning. Additionally, in a home setting, the user can also get distracted, talk with others, etc., during meal preparation time. Everything spoken might not be directed as input to the robot. Therefore, the system should be able to distinguish between conversations with others and valid instructions to the robot and understand the intent of relevant instructions. The system architecture classifies the text into relevant instructions for the robot and understands its intent, which is discussed in the upcoming sections.

3 USER STUDY

A user study is conducted in order to understand how humans would instruct a robot on the necessary actions in order to complete a collaborative cooking task. Figure 1 shows the experimental setup, which consists of two tables. The user sits at the table and instructs the robot to fetch the ingredients and kitchen utensils that are located on the second table. A user study was conducted with 25 adult participants (11 female and 14 male with mean age of 35.4 and standard deviation of 13.1) with the approval of the Institutional Review Board (IRB). The participants are asked to teach the robot how to cook their favorite meal and are instructed to say "Hey robot" and then the action(-s) they would expect the robot to take. However, the participants are not instructed on how to provide the desired robot action. During the study, the robot is responsible for actions related to finding and bringing objects to the participants while the participants would cut the ingredients and cook. Moreover, as we are interested in data collection that is realistic and similar to a home environment, one of the study personnel interrupts and talks with the participant during an instruction or sounds are playing in the background (e.g., dog barking). Participants are also allowed to answer their phone calls and make phone calls.

During the user study, audio data were collected. While collecting data, the majority had their favorite recipe, and depending on the utensils and ingredients that they typically use at their residence, they used specific names for the utensils and ingredients. For example, participants that chose to make pasta each had a unique way of interacting with the robot. The system has no intention of modifying the food-prepping or cooking steps. Instead, it is about creating a comfortable and safe environment by helping users create their meals.

4 PROPOSED SYSTEM

In this work, a system that enables a robot to learn from spoken language is presented. Figure 2 shows the proposed system, which consists of two modules: Natural Language Understanding and Robot Learning.

Natural Language Understanding (NLU): The task of the NLU module is to understand the user's intent. The audio data of the user's speech is converted into text using an offline transcription model, Vosk [22], which can be run locally on the servers to protect the user's privacy to ensure IRB guidelines. The Vosk English language model called the "Vosk-model-en-us-daanzu-20200905" [22] was chosen since it was trained to transcript speech from both native and non-native English speakers. The input to this block is the text transcribed from the user's speech. There is a large likelihood that several interactions will be unrelated to meal preparation during a cooking scenario. The user may get interrupted or distracted by other people and start a conversation with them or answer their phones. Therefore, it is crucial that the system can distinguish between instructions directed toward the robot and unrelated conversations. The sentences that are considered an instruction to the robot are referred to as "valid instruction". To identify and classify valid instructions from other conversations which are considered as "invalid instructions" BERT [11] is used, which is an open-source Bidirectional Encoder Representation from Transformers created by google in 2018. Language models, such as BERT, require a large amount of data to train. Hence, the collected valid instruction data from 25 participants, as previously discussed in Section 3, are randomly augmented by adding food adjectives [23], recipe names [24], and recipe ingredients [25].

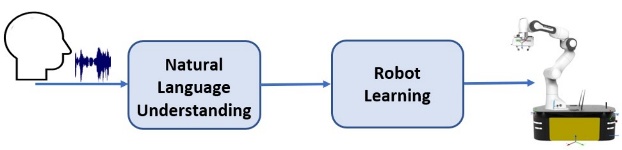

Figure 2: Proposed System for Robot Learning from Spoken Language

The augmented dataset is called the *"Collaborative Cooking dataset"*. The data collected during interruptions and distractions are a small sample from the 25 participants and are not large enough to train the model as examples of 'invalid instruction' for the robot. However, these examples of casual/distracted conversations can be obtained from other datasets, such as TweetQA [26]. The TweetQA is a dataset of informal human conversations and contains 13,757 question-and-answer pairs sampled from 17,794 tweets. Out of which, 10,692 are used for training, and 3,065 are for testing. The tweets from the TweetQA dataset are used to train our BERT model as examples of "invalid instructions." The BERT is used to train to classify between valid and invalid instructions using the Binary Cross Entropy Loss function (L_{BERT}) defined as follows:

$$L_{BERT} = -\sum_{k=0}^{k=n} y_k log(p_k) \tag{1}$$

where p_k is the output class of the model, y_k is the target class, and n is the total number of instructions, which is equal to 20,692 (10,000 from the *Collaborative Cooking dataset* and 10,692 from TweetQA). After an instruction is classified as valid, it is sent to the robot learning module.

Robot Learning: The robot learning module can perform three possible scenarios; (1) new meal preparation, in which the user teaches the robot steps of a new meal, and the robot learns the sequence of the order, (2) meal execution, in which the user can say the name of the learned meal, and the robot would know and execute the sequences, and (3) the user does not care about teaching the robot the entire sequence of the meal preparation; the user is simply requesting a single robot action.

The input of the robot learning module is the valid instructions from the natural language understanding module. The robot learning framework is responsible for building a graph based on the instructed actions for the cooking task. The language model Distilled Generative Pre-trained Transformer-2 (DistilGPT2) [27] generates robot actions from valid instructions. For example, consider a valid instruction S. S needs to be tokenized before giving it as input to the DistilGPT2 model. Tokenization is a process where a sentence is broken down into words, and the words are converted into their unique numerical IDs, which is called the class of the word. The DistilGPT2 model outputs the robot action instruction in tokenized form. To train the DistilGPT2 model, the following loss function $L_{DistilGPT2}$ is used:

$$L_{DistilGPT2} = -\sum_{j=1}^{j=n}\sum_{k=1}^{k=m} T_{jk} log(R_{jk}) \tag{2}$$

where T_{jk} is the target class of jth robot action's k^{th} word (ground truth), R_{jk} is the predicted class of jth robot action's k^{th} word, m is the maximum length of the output robot action that could be generated, n is the total number of valid instruction used to train the DistilGPT2 model, which is 10,000 obtained from the *Collaborative Cooking dataset*. The robot action can be considered a special instruction that captures the intent of the task the user instructs the robot to perform ("execute"). The generated graph consists of the sequence of actions for the specific learned task, as shown in Figure 3. For example, if the user says "Hey robot, bring me the pan", the valid instruction would be generated as "Add to current Graph: Fetch pan". This robot action would be interpreted as adding a node called "Fetch pan" to the graph. After the user has taught the robot how to prepare a meal, the graph is stored in a graph database, which means that the robot learned the sequence of actions for the particular meal.

5 PRELIMINARY RESULTS

To evaluate the complete system proposed in Section 4, we used the data from the *Collaborative Cooking dataset* (see Section 3). The input of our system is the complete speech of the participant for a task, and the output of our system is the learned graph-based sequence of robot actions. The accuracy of the system is defined using the total number (abbreviated as "No.") of correct robot actions generated as shown in Eq. 3.

Task: Denver Omelet

```
Participant: Hey robot today I will be making
denver omelette and I need a saucepan, olive oil
and eggs
DistilGPT-2 Output: Add to Graph: denver omelette
carbonara, Step-1 Fetch Saucepan, Step-2 Fetch
olive oil, Step-3 Fetch eggs

Participant: Hey robot I need ham onion
DistilGPT-2 Output: Add to current graph: Fetch
ham Add to current graph: Fetch onion

Participant's Friend: Hey XYZ omelet smells good I
am famished
system: Invalid command. Text not sent to
DistilGPT-2 for processing

Participant: Hey robot set a timer for 2 minutes
for the eggs to cook
DistilGPT-2 Output: Add to current graph: Set
Timer 10 minutes

Participant: Hey robot Thank you robot for the
help we are finished
DistilGPT-2 Output: Done
```

Corresponding Graph:

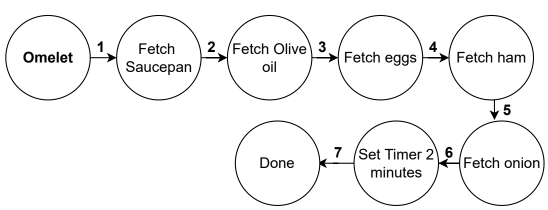

Figure 3: Transcript of the speech given by a participant during the study and the corresponding generated graph for the sequence of robot actions.

$$Accuracy = \frac{\textit{No. of correct actions generated from speech instructions}}{\textit{No. of speech instructions}} \times 100\% \quad (3)$$

Another metric to evaluate the text generated from the DistilGPT2 model is called the BiLingual Evaluation Understudy score (BLEU score for short), which represents how close the generated instruction from the model is to the ground truth instruction. BLEU score ranges from 0 to 1, with 0 meaning no match between the generated and target sequence of actions to 1 being an exact match between the generated and target sequence of actions.

To train the system, the augmented "Collaborative Cooking dataset" from 16 participants is used, and to test the system, the data from the remaining 9 participants is used. The accuracy and BLEU score of the system depends on the training and testing data that are used. For example, some participants used simple sentences like "bring me the pan", while some other participants used comparatively more complex sentences, such as "get that green pan and put it here". When the system is trained using data from participants who used simpler sentences and testing it on the data from participants who used complex sentences would result in less accuracy and BLEU scores and vice versa. Hence, to properly evaluate the system's performance, 3-fold cross-validation is used. The data is divided into folds with respect to the participants that it originated from. The data from 16 participants is used to train the model, and the data from the remaining 9 participants is used to test the model. Finally, the average accuracy and average BLEU score are calculated from different iterations during cross-validation.

The average accuracy calculated is 92.85 ± 3.87%, and the average BLEU score calculated is 0.91 ± 0.11. There were some cases where the system was not able to generate an action from a spoken instruction. For example, one of the participants said 'I feel like eating pizza". This instruction was classified as invalid because it is ambiguous whether this sentence is a casual conversation or a valid instruction. In some cases, the DistilGPT2 was not able to generate a correct robot action instruction, e.g. "Also I need some glass baking dish". The "glass baking dish" was identified as a task/recipe name rather than an object that needs to be fetched.

6 CONCLUSION & FUTURE WORK

In this paper, we introduced a robot learning framework that enables the robot to learn from unstructured spoken language. A collaborative meal preparation scenario between individuals and the robot is developed, and data were collected. Individuals were comfortable communicating verbally with the robot in order to get assistance with the meal preparations in the kitchen. We propose a framework that automatically generates robot actions from speech that included environmental interruptions. A 3-fold cross-validation average accuracy and average BLEU score were calculated. The average accuracy of the system is 92.85 ± 3.87%, and the average BLEU score of the system is 0.91 ± 0.11, which is a very promising and positive finding. The prospect of natural interaction using speech opens up many possibilities that can be further explored.

In the future, the graph-based sequence of actions will be connected to a robot skill library, similar to [28], that would enable the robot to perform the actions in real-time. During the data collection, we noticed that some participants felt comfortable interacting with the robot, and they leaned on using more fragments phrases. For example, one of the participants did not remember how to request the "wooden spoon", and instead, they said "give me the wooden thingy". In another case, the participant mentioned "stainer" instead of "strainer". Interestingly enough, we googled the word "stainer", and the top search items were "strainers". Therefore, we plan to have the user teach the robot the name of the items the way they prefer to refer to them. Another observation was that several participants used hand gestures while asking for an item. In the future, we plan to conduct a larger study to ensure our proposed framework is robust and publish the dataset.

7 ACKNOWLEDGMENT

We extend our gratitude to the participants for their invaluable contributions to our user study. Additionally, this material is based upon work supported partially supported by the National Science Foundation under Grant 2226165 and by Santa Clara University under University Research Grant.

Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

[1] CDC, "Disability impacts all of us," 2022. [Online]. Available: https://www.cdc.gov/ncbddd/disabilityandhealth/infographic-disability-impacts-all.html#:~:text=61%20million%20adults%20in%20the,have%20some%20type%20of%20disability

[2] WHO, "Health and ageing," 2022. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/ageing-and-health#:~:text=By%202030%2C%201%20in%206,will%20double%20(2.1%20billion)

[3] C.-S. Chung, H. Wang, and R. A. Cooper, "Functional assessment and performance evaluation for assistive robotic manipulators: Literature review," *The journal of spinal cord medicine*, vol. 36, no. 4, pp. 273–289, 2013.

[4] W.-T. Ma, W.-X. Yan, Z. Fu, and Y.-Z. Zhao, "A chinese cooking robot for elderly and disabled people," *Robotica*, vol. 29, no. 6, pp. 843–852, 2011.

[5] M. Kyrarini, Q. Zheng, M. A. Haseeb, and A. Gräser, "Robot learning of assistive manipulation tasks by demonstration via head gesture-based interface," in *2019 IEEE 16th International Conference on Rehabilitation Robotics (ICORR)*. IEEE, 2019, pp. 1139–1146.

[6] F. F. Goldau, T. K. Shastha, M. Kyrarini, and A. Gräser, "Autonomous multi-sensory robotic assistant for a drinking task," in *2019 IEEE 16th International Conference on Rehabilitation Robotics (ICORR)*. IEEE, 2019, pp. 210–216.

[7] K. Kronhardt, S. Rübner, M. Pascher, F. F. Goldau, U. Frese, and J. Gerken, "Adapt or perish? exploring the effectiveness of adaptive dof control interaction methods for assistive robot arms," *Technologies*, vol. 10, no. 1, p. 30, 2022.

[8] M. Biswas, M. Romeo, A. Cangelosi, and R. B. Jones, "Are older people any different from younger people in the way they want to interact with robots? scenario based survey," *Journal on Multimodal User Interfaces*, vol. 14, pp. 61–72, 3 2020. [Online]. Available: https://link.springer.com/article/10.1007/s12193-019-00306-x

[9] M. Marge, C. Espy-Wilson, N. G. Ward, A. Alwan, Y. Artzi, M. Bansal, G. Blankenship, J. Chai, H. Daumé III, D. Dey *et al.*, "Spoken language interaction with robots: Recommendations for future research," *Computer Speech & Language*, vol. 71, p. 101255, 2022.

[10] L. Shao, T. Migimatsu, Q. Zhang, K. Yang, and J. Bohg, "Concept2robot: Learning manipulation concepts from instructions and human demonstrations," in *Proceedings of Robotics: Science and Systems (RSS)*, 2020.

[11] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," 2018. [Online]. Available: https://arxiv.org/abs/1810.04805

[12] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015. [Online]. Available: https://arxiv.org/abs/1512.03385

[13] R. Goyal, S. E. Kahou, V. Michalski, J. Materzyńska, S. Westphal, H. Kim, V. Haenel, I. Fruend, P. Yianilos, M. Mueller-Freitag, F. Hoppe, C. Thurau, I. Bax, and R. Memisevic, "The "something something" video database for learning and evaluating visual common sense," 2017. [Online]. Available: https://arxiv.org/abs/1706.04261

[14] I. Giorgi, A. Cangelosi, and G. L. Masala, "Learning actions from natural language instructions using an on-world embodied cognitive architecture," *Frontiers in Neurorobotics*, vol. 15, p. 48, 5 2021.

[15] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," 2015. [Online]. Available: https://arxiv.org/abs/1506.02640

[16] V. V. Unhelkar, S. Li, and J. A. Shah, "Decision-making for bidirectional communication in sequential human-robot collaborative tasks," *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*. [Online]. Available: https://doi.org/10.1145/3319502.3374779

[17] D. Choi, W. Shi, Y. S. Liang, K. H. Yeo, and J. J. Kim, "Controlling industrial robots with high-level verbal commands," *Lecture Notes in Computer Science including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics*, vol. 13086 LNAI, pp. 216–226, 2021. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-030-90525-5_19

[18] C. Li, X. Zhang, D. Chrysostomou, and H. Yang, "Tod4ir: A humanised task-oriented dialogue system for industrial robots," *IEEE Access*, pp. 1–1, 8 2022.

[19] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, "Language models are unsupervised multitask learners," 2018. [Online]. Available: https://d4mucfpksywv.cloudfront.net/better-language-models/language-models.pdf

[20] M. Ahn, A. Brohan, N. Brown, Y. Chebotar, O. Cortes, B. David, C. Finn, C. Fu, K. Gopalakrishnan, K. Hausman, A. Herzog, D. Ho, J. Hsu, J. Ibarz, B. Ichter, A. Irpan, E. Jang, R. J. Ruano, K. Jeffrey, S. Jesmonth, N. J. Joshi, R. Julian, D. Kalashnikov, Y. Kuang, K.-H. Lee, S. Levine, Y. Lu, L. Luu, C. Parada, P. Pastor, J. Quiambao, K. Rao, J. Rettinghouse, D. Reyes, P. Sermanet, N. Sievers, C. Tan, A. Toshev, V. Vanhoucke, F. Xia, T. Xiao, P. Xu, S. Xu, M. Yan, and A. Zeng, "Do as i can, not as i say: Grounding language in robotic affordances," 2022. [Online]. Available: https://arxiv.org/abs/2204.01691

[21] S. Kuoppamäki, S. Tuncer, S. Eriksson, and D. McMillan, "Designing kitchen technologies for ageing in place," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 5, p. 19, 6 2021. [Online]. Available: https://dl.acm.org/doi/10.1145/3463516

[22] "Vosk models." [Online]. Available: https://alphacephei.com/vosk/models

[23] EnglishStudyHere, "Food adjectives – list of food adjectives," 2018. [Online]. Available: https://englishstudyhere.com/grammar/adjectives/food-adjectives-list-of-food-adjectives/

[24] F. Network, "Recipes a to z," 2022. [Online]. Available: https://www.foodnetwork.com/recipes/recipes-a-z/123

[25] O. S. University, "Ingredients list," 2021. [Online]. Available: https://www.foodhero.org/ingredients

[26] W. Xiong, J. Wu, H. Wang, V. Kulkarni, M. Yu, S. Chang, X. Guo, and W. Y. Wang, "Tweetqa: A social media focused question answering dataset," 2019. [Online]. Available: https://arxiv.org/abs/1907.06292

[27] "Distilgpt2." [Online]. Available: https://huggingface.co/distilgpt2

[28] M. Kyrarini, S. Naeem, X. Wang, and A. Gräser, "Skill robot library: Intelligent path planning framework for object manipulation," in *2017 25th European Signal Processing Conference (EUSIPCO)*. IEEE, 2017, pp. 2398–2402.