Wait Wait, Nonverbally Tell Me: Legibility for Use in Restaurant Navigation

Ada V. Taylor
Robotics Institute,
Carnegie Mellon University
Email: adat+rss@andrew.cmu.edu

Ellie Mamantov
Computer Science
Yale University
Email: ellie.mamantov@yale.edu

Henny Admoni
Robotics Institute,
Carnegie Mellon University
Email: hadmoni@andrew.cmu.edu

***Abstract*—Legible motion is important for robots navigating in social environments, but certain contexts present additional constraints that require us to adapt what actions count as legible. In this work, we explore the legibility of navigation paths used by a robot waiter in a restaurant context. We highlight several areas where the existing formulation needs expansion for this context, with three areas of potential improvement: distribution of highly legible moments to within the field of view of observers, balancing between multiple audience members with different perspectives, and incorporating other social signals relevant to goal inference. We design an experimental scenario to begin examining the effectiveness of these models of improved "audience-aware" legibility.**

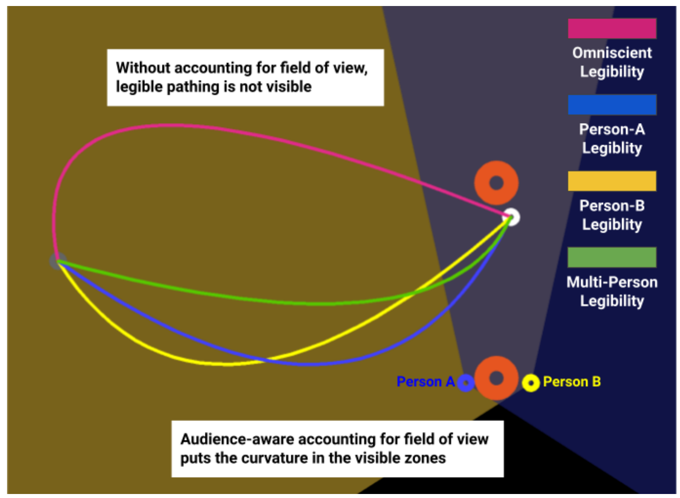

Fig. 1. An overview of differently-legible paths in a restaurant environment with four tables, and the robot starting at the left. The field of view of Person A can be seen in blue extending towards the right, and that of Person B in yellow extending to the left. The top path in pink uses the original legibility formulation which assumes complete or "omniscient" vision of the restaurant. The bottom paths are generated with an "audience-aware" model of the limited perspectives of each of the customers in their corresponding colors, and with their combined perspectives being used to generate the middle path in green.

I. Introduction

As robots begin to be used in increasingly human-dense environments, they need to be able to express their intent in unobtrusive, nonverbal ways that indicate their goals to observers. In many situations, it is valuable for a robot to be able to convey its intent implicitly [2], and being able to implicitly communicate goals can allow for more fluent and efficient human-robot collaboration in a shared space [11].

We are motivated by the scenario of a robot waiter visiting a table at a restaurant to provide service. The goal is to generate approach trajectories for a robot server that are more human-understandable and informative to the customers in the restaurant, as measured by their correct understanding of the goal of a robot over time and thus readiness for its arrival.

While currently limited in use, robots have already been deployed in commercial restaurant settings [14]. Among those deployed for serving tasks, the majority of their utility comes from traveling to and from tables: taking orders [13], transporting food to aid a human waiter [4], or interacting with guests [12].

An effective legible approach here could enable more effective service by dynamically informing path planning, or simply optimizing these static paths for better customer experience.

A. Unique Aspects of Restaurant Environment

Legibility in a restaurant environment has some unique factors as compared to end effector movements, due to the nature of the environment. These include:

1) Limited Field of View: While the concepts of legibility can be deployed regardless of scale, the larger restaurant environment requires longer movements relative to the observer. Therefore the robot is likely not observable for the entire path. Additionally, unlike in a scenario with a robot performing a grasping task, the human observer is likely to be oriented towards their fellow tablemate or meal, and tracking the robot's movements is secondary. Both of these indicate that the robot is likely to go in and out of view, and that modeling the observer's perspective is important for this problem.

2) Multiple Audience Members: As the robot traverses through this larger environment, they also go in and out of the vision of multiple observers. This means the robot needs to account for this entire audience, even though they may all have different perspectives of the scene.

3) Social Interaction: Because this is a social space, and the robot is humanoid, it may be that other social signals dominate over an assessment of legibility based purely on functional signals such as movements.

These factors are broadly relevant to legibility as applied to social navigation scenarios.

B. Intent-Expressive Motion

Legible robot motion is path planning in a manner that clarifies the robot's objective in order to support human interaction. These motions are designed to allow a human observer to infer the robot's intent more confidently and quickly [1, 7, 8, 9]. It has been shown that path trajectories that do not match humans' expectations but convey the motion's goal or intent are less predictable but more legible, and allow a human observer to infer the robot's intent more confidently and quickly [3]. Legibility has been extended to generate socially communicative signals such as pointing, in addition to movements in space [5].

To develop restaurant-specific legible navigation, we have created a formulation of legibility based on Dragan et al's work [3, 10] that takes into account the relative visibility of the robot with multiple observers' specific fields of view in mind. This legibility is used to select paths to each possible goal that have consistent cost from an omniscient point of view, but variable legibility according to our audience-aware algorithm. We detail this in section II.

Most contemporary legible motion planners assume that human observers are omniscient and are aware of all changes in the robot's configuration as it moves, even though human observers have a particular field of view that does not encompass the entire scene. One previous study investigated this assumption and introduced a model that optimizes motions with a certain 2D-projection in mind; it was shown that certain viewing angles can lead to depth uncertainty and occlusions that make certain trajectories less legible [10]. However, this technique does not deal with the overall field of view or multiple perspectives, but instead focuses on how motions within a field of view are perceived.

II. Legibility Formulation

In order to mathematically account for these important aspects of audience-aware legibility, we take as a starting point the legibility equation created by Dragan et al in [3, 10]. This equation is built around a model of how the user assesses the likelihood of approaching each goal location given a particular path ξ, with the goal of maximizing the clarity of that inference over the entire path to create a highly legible, intent-expressive route.

$$\text{legibility}(\xi) = \frac{\int P(G^*|\xi_{S\rightarrow\xi(t)}) f(t) dt}{\int f(t)} \tag{1}$$

This inference of $P(G^*|\xi_{S\rightarrow\xi(t)})$ is based on the efficiency of movement along the path so far compared to the relative locations of all the goals in set $\mathcal{G}$ from the current point.

III. Adapting Legibility to a Restaurant Context

A. Modeling Limited Fields of View

In this original formulation of legibility, the distribution of highly legible moments is weighted towards the beginning of the path with the function $f(t)$, which is by default $f(t) = T - t$, a linearly decreasing function from the beginning of the time period to the end. This is intended to incentivize early highly-legible moments that enable the observer to infer the correct goal as quickly as possible.

However, in a restaurant context, motions tend to take place over a much broader space than observers can see. Early movements on the path are in fact likely to be farthest away and thus often outside of their field of view. Therefore, we need to incorporate a representation of visibility into the $f(t)$.

Given the angle of an observer in the environment as θ_O, and the angle between observer's location to a given point at $\xi(t)$ as $\theta_{O,\xi(t)}$, we incentivize movements centered within the user's field of view θ_{FOV} by defining visibility like so:

$$V(\xi(t), O) = \begin{cases} |\theta_{FOV} - (\theta_O - \theta_{O,\xi(t))})| \\ \qquad\qquad \text{when } (\theta_O - \theta_{O,\xi(t)}) < \theta_{FOV} \\ 0 \qquad\quad \text{otherwise} \end{cases} \tag{2}$$

Intuitively, points along the path which are invisible to the observer will have a visibility of 0, and once inside the viewable range, visibility linearly increases as an object gets closer to the center of an observer's field of vision. To integrate this into the original legibility equation, we can use $V(\xi(t), O)$ as our function $f(t)$.

Now, unseen movements will no longer contribute the total quantification of legibility, and movements closer to the center of vision are preferred.

B. Balancing Multiple Observers

Once we recognize that each observer has a limited area of vision, we then understand that we may need to need create a representation of visibility that can account for multiple unique observers at once. In a restaurant context it is very common that not all observers have equivalent views of the working area within a scene.

Notably, observers with "omniscient" perfect knowledge of the path or a different limited perspective may find a path customized for another individual to be confusing; this is similar to viewing a magic trick from an unexpected angle. We expect the effectiveness of the communication to be dependent on modeling each observer's perspective.

The simplest method of accounting for all observers is to sum their visibilities:

$$f(t) = V_{sum}(\xi(t), \mathcal{O}) = \sum_{O\in\mathcal{O}} V(\xi(t), O) \tag{3}$$

However, while a simple summation is a good first step, it is insufficient for this problem. Our goal is not to simply make a maximally visible path. For a table-waiting context, we actually want to maximize the window for which both relevant members are ready for the arrival of the robot as long as possible. In equation 3, paths visible to one customer but out of the view of the other can easily dominate over those with a shorter but more useful shared window of readiness.

To illustrate this asymmetry, we can see in figure 2 the perspectives of two customers in both 2D and 3D space. When

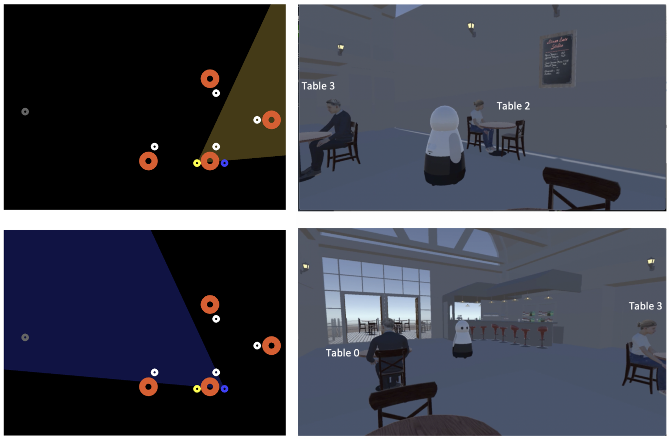

Fig. 2. User perspectives of the scene from each audience position in the restaurant, both in 2D pathing simulation and 3D restaurant simulation. Four tables in orange are represented alongside their corresponding goal locations in smaller white circles. Person A is shown in the top row, with Person B's view in the second.

a robot approaches the bottom table from the left, the customer on the left has their back to the approach, leading to a much more difficult scenario for being able to observe the approach of the robot. On the other hand, the customer to the right is able to watch the entire path, and is much more likely to have a long window of readiness.

In a more general sense, we would like to maximize legibility for observers during the period for which the robot's movements intersect with their goals. For a group of seated customers, this is a large window of clearly understanding the robot's goal table. We can modify the function $f(t)$ like so:

$$V_{min}(\xi(t), \mathcal{O}) = \min_{\forall O \in \mathcal{O}} V(\xi(t), O) \tag{4}$$

This definition enables the most disadvantaged viewpoint to have the best chance possible of seeing the path. However, as additional audience members with varied fields of view are added, the zone for which they can all understand the scene becomes limited. This definition may also have drawbacks when a majority of observers have a similar view, but one outlier with a poor viewpoint dominates the multiview equation.

Therefore, it would perhaps be more efficient and effective for a robot to choose what subset of restaurant members is relevant for its legibility as it traverses a space. Group member determination might be made based on a combination of each individual observer's perspective of the space and what communication goals they are watching for. For example, we may only be interested in members of the table the robot is heading to being able to understand our path.

We may also find that attempting to actively exclude the customers who are not meant to receive our nonverbal "message" when composing paths produces simpler paths with tighter customization, and avoid misleading non-target customers. It may be useful to avoid their fields of vision as much as possible.

Understanding the correct audience scope could also help the robot switch between highly-efficient and highly-expressive policies at the relevant junctures.

C. Social Aspects of Intention Expressiveness

Humans have different expectations of a humanoid robot than the tip of an end effector, particularly when a head, eyes, or audio cues are involved [6]. While our previous formulations relied on a definition of the probability for a given goal G^* in terms of how efficient movement was towards that goal as compared to the other goals, it may be that this signal does not dominate when humans make assessments of a humanoid robot.

Notably, our earlier paradigms of degree-of-observability and balancing the visibility from multiple perspectives are applicable to any intent-expressive behavior. Even discretized behaviors such as signals or displays that directly indicate which goal is the target [15] could be expressed. Using the example of a robot which displays the table number of its current goal location, we can integrate it like so:

$$P(G|\text{table \#}) = \delta(\text{table \#} = \text{mine}) \tag{5}$$

Here, $\delta(a = b)$ is the Kronecker delta function, which evaluates to 1 if the arguments are equal and 0 otherwise.

We could also imagine a description of robot heading as the indicator of intent, with the probability based on the angle between the current robot heading and a goal $\theta_{robot,G}$ at time t along path ξ:

$$P(G^*|\xi, t) = \frac{|\theta_{(robot,G^*)}(t)|}{\sum_{G\in\mathbf{G}} |\theta_{(robot,G)}(t)|} \tag{6}$$

In the case where the robot has a separately articulated head or the robot is capable of omnidirectional motion, this can be set separately from the heading. If the "head" is static and fixed to the chassis, as in the case of an RC car, then $\theta_{robot}(t)$ is dependent on $\xi_{S\rightarrow\xi(t)}$.

IV. Current Work: Audience-Aware Legibility

We plan to test the effect of taking into account audience viewpoints when planning approach trajectories through a video-based human subjects experiment. Our study compares the legibility of trajectories that were personalized for unique or combined perspectives to a baseline of an omniscient perspective.

Hypotheses: We explore an inherent strength and weakness to customizing paths to a particular audience:

- **H1** A robot that plans its path taking into account the visibility of that path to all of its observers can create paths which are easier for viewers to understand.
- **H2** A path personalized for a specific perspective will be more legible than the average multi-perspective performance for that perspective but will be less legible for other perspectives.

Intuitively, leveraging more information could allow a robot to accommodate its audience more effectively. The drawback of this specificity is that groups who are not being catered to do not benefit, and can potentially be confused by seemingly unnecessary or incorrectly evocative robot movement.

We plan to test the effect of accounting for these perspectives through an online human subjects experiment which records user expectations of goal across the entire path.

A. Experimental Procedures:

To investigate if this approach clarifies intent for the observer, we have designed the following experiment:

Task: Participants were shown videos of a robot server approaching different "goal tables" in the restaurant. They were tasked with indicating how confident they are that the server is approaching their table.

Stimuli: We planned our video stimuli by manipulating several variables. In every video, a robot moves from a starting location to one of four goal tables: the participant's table or a table in front of, across from, or perpendicular to the participant's table. Given this basic setup, we manipulate:

> ***Perspective,*** by changing which of the chairs at the participant's table they are watching the scene from;
>
> ***Path Planning Method***, by having the robot follow a path planned for an omniscient perspective, both perspectives, or each single perspective.

B. Metrics

Our metrics focus on quantifying user certainty and correctness over the interval of the trial, as well as a moment when they are most sure.

We understand that there will be a zone where users are uncertain about their choice, or small offsets can occur, so we have divided slider responses into three zones: *uncertain*, *correct*, and *incorrect*. *Uncertain* is defined as +/- 5% from the middle of the slider, *correct* is above that threshold with the correct polarity, and *incorrect* is below that threshold with the wrong polarity.

We also want to know the final window for which users were correct about the robot's goal, which is defined as the length of the final period for which a user is *correct*. This *envelope of certainty* can also be calculated for a higher threshold, or simply above/below neutral. Finally, we also account for times that the path is actively misleading or confusing, counting the number of "reversals" as times when the user's assessment completely crossed the zone of uncertainty from either *correct* to *incorrect*, or vice-versa.

C. Deployment

We have developed a 3D restaurant simulator with a robot server, and we replay each of our 2D paths in this 3D environment, captured from the perspective of each audience location. During the trials, participants use a continuous slider to report how confident they are that the server is approaching their table. The post-study questionnaire includes open-ended questions for participants to provide feedback.

The study will be deployed online, and consists of a tutorial that introduces the participants to the restaurant's layout and the mechanics of the slider that they will use to report their confidence, the trials, and a final questionnaire. To ensure that no participant sees the same path twice (but from different perspectives), each participant is randomly assigned to one of the two perspectives. Participants report their confidence continuously, and the video will only continue playing if they were actively holding the slider.

A pilot study indicated that path visibility does indeed improve the interval for which users were certain of the robot's goal, even before taking legibility into account.

V. Future Work

Our work provides an initial method for legibility in contexts with a limited field of view, but additional questions remain. How does this model of limited viewpoints improve audience reactions to paths? We expect to see increased windows where both customers are ready for the approach of the waiter. As we review these results, we hope to explore the relationship between the offset between viewpoints and the degree to which their envelope of certainty decays.

We also plan to explore whether our model of vision is appropriate by directly asking users about their inferences and searching for a relationship between the centrality of motion within the user's vision and their certainty. It may be that there is no benefit to incentivizing motion in the center of vision as long as it is at all within view.

With a solid model of field of view, we hope to explore our questions around balancing audiences. We have provided a few preliminary examples of multi-view incorporation into legibility in equations 3 and 4, but what are the best representations of multi-viewer legibility across different environments? To explore this, we need to research scenarios with different audience members across a variety of similar and dissimilar viewpoints.

We hope that this exploration will give us insight into the question of when a particular customer is appropriate to add to our scope. As we investigate the limitations of catering to an increasingly larger audience, we can explore whether paths are most effective when catering to all possible perspectives, or a few relevant to the target task. Additionally, in a busy environment with multiple audience subsections, does it increase overall clarity to actively exclude and avoid the observable areas of customers who are not the target of the interaction, therefore avoiding misleading customers who should not be expecting a visit from the robot?

Finally, we hope to blend our motion-based assessments of legibility with other social signals of intention found in a restaurant scenario. Given that heading, gaze direction, or other overt signals are possible, which models of how observers infer the probability of approaching a particular goal are most relevant in a social navigation scenario? And how do the dominant cues relate to the robot's design and elements such as a head or eyes? We expect that the more human-like a robot is, the more powerful its cues outside of movement can be in providing information about its intentions.

ACKNOWLEDGMENTS

This work is supported in part by the National Science Foundation (#IIS1755823, #DGE1745016) and the Sony Group Corporation.

REFERENCES

[1] Michael Beetz, Freek Stulp, Piotr Esden-Tempski, Andreas Fedrizzi, Ulrich Klank, Ingo Kresse, Alexis Maldonado, and Federico Ruiz. Generality and legibility in mobile manipulation. *Autonomous Robots*, 28(1):21, 2010.

[2] Elizabeth Cha, Yunkyung Kim, Terrence Fong, Maja J Mataric, et al. A survey of nonverbal signaling methods for non-humanoid robots. *Foundations and Trends® in Robotics*, 6(4):211–323, 2018.

[3] Anca D Dragan, Kenton CT Lee, and Siddhartha S Srinivasa. Legibility and predictability of robot motion. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 301–308. IEEE, 2013.

[4] Pantelis Elinas, Jesse Hoey, Darrell Lahey, Jefferson D Montgomery, Don Murray, Stephen Se, and James J Little. Waiting with josé, a vision-based mobile robot. In *Proceedings 2002 IEEE International Conference on Robotics and Automation (Cat. No. 02CH37292)*, volume 4, pages 3698–3705. IEEE, 2002.

[5] Rachel M Holladay, Anca D Dragan, and Siddhartha S Srinivasa. Legible robot pointing. In *The 23rd IEEE International Symposium on robot and human interactive communication*, pages 217–223. IEEE, 2014.

[6] Heather Knight and Reid Simmons. Laban head-motions convey robot state: A call for robot body language. In *2016 IEEE international conference on robotics and automation (ICRA)*, pages 2881–2888. IEEE, 2016.

[7] Christina Lichtenthäler and Alexandra Kirsch. Towards legible robot navigation-how to increase the intend expressiveness of robot navigation behavior. 2013.

[8] Christina Lichtenthäler and Alexandra Kirsch. Goal-predictability vs. trajectory-predictability: which legibility factor counts. In *Proceedings of the 2014 acm/ieee international conference on human-robot interaction*, pages 228–229, 2014.

[9] Christina Lichtenthäler and Alexandra Kirsch. Legibility of robot behavior: a literature review. 2016.

[10] Stefanos Nikolaidis, Anca Dragan, and Siddhartha Srinivasa. Based legibility optimization. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 271–278. IEEE, 2016.

[11] Stefania Pellegrinelli, Henny Admoni, Shervin Javdani, and Siddhartha Srinivasa. Human-robot shared workspace collaboration via hindsight optimization. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 831–838. IEEE, 2016.

[12] Ronald PA Petrick and Mary Ellen Foster. What would you like to drink? recognising and planning with social states in a robot bartender domain. In *CogRob@ AAAI*, 2012.

[13] Sakari Pieskä, Markus Liuska, Juhana Jauhiainen, Antti Auno, and Delitaz Oy. Intelligent restaurant system smartmenu. In *2013 IEEE 4th International Conference on Cognitive Infocommunications (CogInfoCom)*, pages 625–630. IEEE, 2013.

[14] Sakari Pieska, Mika Luimula, Juhana Jauhiainen, and Van Spiz. Social service robots in wellness and restaurant applications. *Journal of Communication and Computer*, 10(1):116–123, 2013.

[15] Thomas Weng, Leah Perlmutter, Stefanos Nikolaidis, Siddhartha Srinivasa, and Maya Cakmak. Robot object referencing through legible situated projections. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 8004–8010. IEEE, 2019.