"Yeess!" "No-oh-oh..." Implicit Robot Task Information from Prosody in Human Verbal Feedback

Murat Han Aydoğan
VU Amsterdam
maydogan17@ku.edu.tr

Kenneth D Mitra
University of Texas at Austin
kennethmitra@utexas.edu

Kush Desai
University of Texas at Austin
kushkdesai@utexas.edu

Taylor Kessler Faulkner
University of Washington
taylorkf@uw.edu

Akanksha Saran*
Microsoft Research
akanksha.saran@microsoft.com

Kim Baraka*
VU Amsterdam
k.baraka@vu.nl

***Abstract*—This paper presents preliminary evidence that prosody carries useful task information in an interactive reinforcement learning setting with verbal evaluative feedback from a human teacher. We developed a novel mixed-participant Wizard-of-Oz study setup to collect audio data from participants teaching a reinforcement learning agent in a grid world navigation task where the agent was wizarded by another participant, hence simulating prosody-sensitive agent learning. Our pilot study shows that, for the participants tested in the teacher role, prosodic features such as energy and pitch are statistically significantly correlated with the advantage function, an underlying Markov Decision Process feature used by RL algorithms. Results also suggest some level of individual differences between different teachers. While further research is needed to develop computational models of human teachers' prosody in different learning tasks, our early results highlight the potential of tapping into implicit voice signals to improve robot learning and human teaching efficiency.**

***Index Terms*—Interactive reinforcement learning, Prosody, Human-agent interaction, Wizard-of-Oz**

I. Introduction

In order to allow users to have more control over a robot's operation, we believe robots should be able to learn from interaction with said users who can take on the role of human teachers. One popular approach to address robot learning from interaction with human teachers is interactive reinforcement learning (intRL), whereby a user provides evaluative feedback (e.g., binary positive/negative feedback) to a learning agent, in order to influence or shape its policy over time [1]–[3]. This setup allows the human teacher to intuitively provide feedback on the robot's actions without requiring demonstrations, which are often difficult for non-expert users to provide. Teachers also have more flexibility in personalizing the robot's behavior to their own preferences. One popular way for communicating this feedback is the use of speech interfaces [4]–[8]. While these interfaces are intuitive for untrained users, they mostly focus on the content of the speech (the words being uttered) and ignore other parts of the voice signals that may be relevant for the learner.

The claim we make in this paper is that *prosody*, as an implicit paralinguistic signal present in human voice, *carries useful task information*. As such, it could be considered as a valid teaching signal that complements speech content in future intRL research. We conducted a mixed-participant Wizard-of-Oz (WoZ) study where, in each session, two participants are secretly matched, one in the role of a teacher giving "yes/no" feedback, and one in the role of the Wizard controlling the agent through a keyboard. This setup allowed us to collect naturalistic teacher data by simulating the learning process of the agent using a human Wizard with limited task information. We then ran a correlation analysis on the teachers' voice data to relate prosodic features to features of the underlying Markov Decision Process (MDP) used to model the learning task.

Our contributions are summarized as follows:

- A novel mixed-participant WoZ experimental setup to collect naturalistic teacher voice data.
- A pilot study showing significant correlations between prosodic features of voice data (namely, energy and pitch) with underlying MPD features (namely, the advantage function).

II. Related Work

We provide a brief overview of work on learning from human audio as well as other implicit social signals.

A. Learning from human audio

Some prior works which leverage human feedback during reinforcement learning tasks [9] do so via voice [4]–[8]. Tenorio et al. [4] perform reward shaping using SARSA [10]–[12] with human voice. Under their setup, the voice-based feedback is provided as the robot is executing the task. However, rewards are predefined for certain words in a vocabulary of 250 words such as +50 for 'excellent', -50 for 'terrible' etc., and no

* Equal contribution

prosodic information is used. Krenig et al. [7] train RL agents with action advice in the form of human voice, such that a set of predefined words directly map to an underlying action from a discrete action set. Krenig et al. [8] use sentiment analysis to filter explanations into advice of what to do and warnings of what to avoid. Nicolescu et al. [13] demonstrated the role of verbal cues both during demonstrations and as feedback from the human teacher during the agent's learning process, to facilitate learning of navigation behaviors on a mobile robot, with a limited vocabulary of words to indicate relevant parts of the workspace or actions that a robot must execute. Similarly, Pardowitz et al. [14] used a fixed set of seven vocal comments which are mapped one-to-one with features relevant to the task to augment subtask similarity detection and learning of the task model from demonstrations for a simple table setting task. However, all of these prior works focus on the spoken words and do not leverage prosody from human speech—an informative and rich signal of human intent which has the potential to further enhance learning [15].

Kim et al. [5] use affective human speech feedback over 25 msec audio snippets to improve a social waving behavior using Q learning. They use three prosody features (total band energy, variance of log-magnitude-spectrum, variance of log-spectral-energy) to learn the wave which optimally satisfies a human tutor. We build on this work to further understand how prosodic features relate to MDP features, such as the advantage function, with the goal to inform the design of future RL algorithms which can be more sample efficient by leveraging prosodic information.

B. Learning from other implicit signals

The field of robot learning with multi-modal human cues [16], [17] has leveraged other implicit signals apart from speech, such as clicker-based feedback (perfect and imperfect) [18]–[20], eye movements [21]–[24], facial expressions [25], [26], gestures [26], [27], haptic feedback [28], [29], object and environmental sounds [30]–[33]. Verbal prosodic feedback is a relatively under-explored modality for intRL. However, there has been some recent advances which highlight the richness of prosody for robot learning from demonstration [15]. In our work, we build on this work to further develop an understanding of human prosodic features for intRL.

III. Mixed-participant Wizard-of-Oz setup

In order to develop an algorithm that leverages implicit information in prosody, we need to first understand how people use prosody as a teaching signal. On the other hand, in order to understand prosodic behavior in this context, we need to have an algorithm that incorporates prosody in its learning, which we don't have yet. To solve this paradox, we opted for a mixed participant Wizard-of-Oz approach where one participant plays the role of a teacher, and the other participant with no information about the task plays the role of a Wizard. This setup makes sure that the teacher audio we get is as close a possible to what we would expect in our target context. This approach is superior to using a baseline algorithm for the agent (e.g., intRL based on speech only) for two reasons. First, we expect the teacher to adapt to the learner, thereby potentially suppressing their prosodic signals (which was confirmed in some early pilots we ran with fixed agent trajectories). Second, the wizard's keystrokes provide us with valuable data that can be used in future research to better understand local and global interpretations of teacher feedback, independently of how well the teacher is able to teach.

Our contributed web-based WoZ interface is shown in Fig. 1. While the teacher sees the full environment and provides online verbal feedback to the agent in real-time, the Wizard receives the teacher audio in real-time (streamed through Twilio, a secure web service) and needs to control the agent through keystrokes in response to current and past feedback from the teacher. The Wizard is shown a sanitized view of the environment that only shows the grid. The two environments are synchronized over web sockets to ensure consistent agent positions on both interfaces.

IV. Study design

A. Participants

We recruited eight university students as participants for our pilot study, four of which were assigned the role of teachers and four of them the role of wizard (randomly), with English fluency as an inclusion criterion. In the teacher role, half the participants identified as female and the other half identified as male. Three of them were native English speakers. They all have experience with playing video games and three of them have experience with programming. On the wizard side, three identified themselves as male and one as female. Two of them were native English speakers. All of the wizards have experience with programming and playing video games.

B. Experimental setup

Figure 1 shows what each participant saw on the screen. The teacher sees the whole map of the game with full details while the wizard only sees the robot and the map. The Wizard had no previous idea about what the task entailed or where the special states were located. The map was created with wall borders around the playable area. The robot location was initialized at the start of the game with a uniform random selection. The robot's starting direction was picked randomly from four options: up, down, left, and right. Also, the robot was initialized in such a way that it could move three spaces in its starting direction without hitting a wall. Game elements were placed last with the following constraints: bombs must not be within 3 spaces of the robot's initial direction of travel and the Manhattan distance between elements must be 4 or greater. Unknown to the teacher, the wizard controlled the robot with arrow keys on the keyboard and in the absence of wizard input for 1.25 seconds or more, the robot started exploring the map randomly, mimicking exploration/exploitation phases of most RL algorithms. The game concluded after one practice round (until the goal was reached) and three actual game rounds used for analysis.

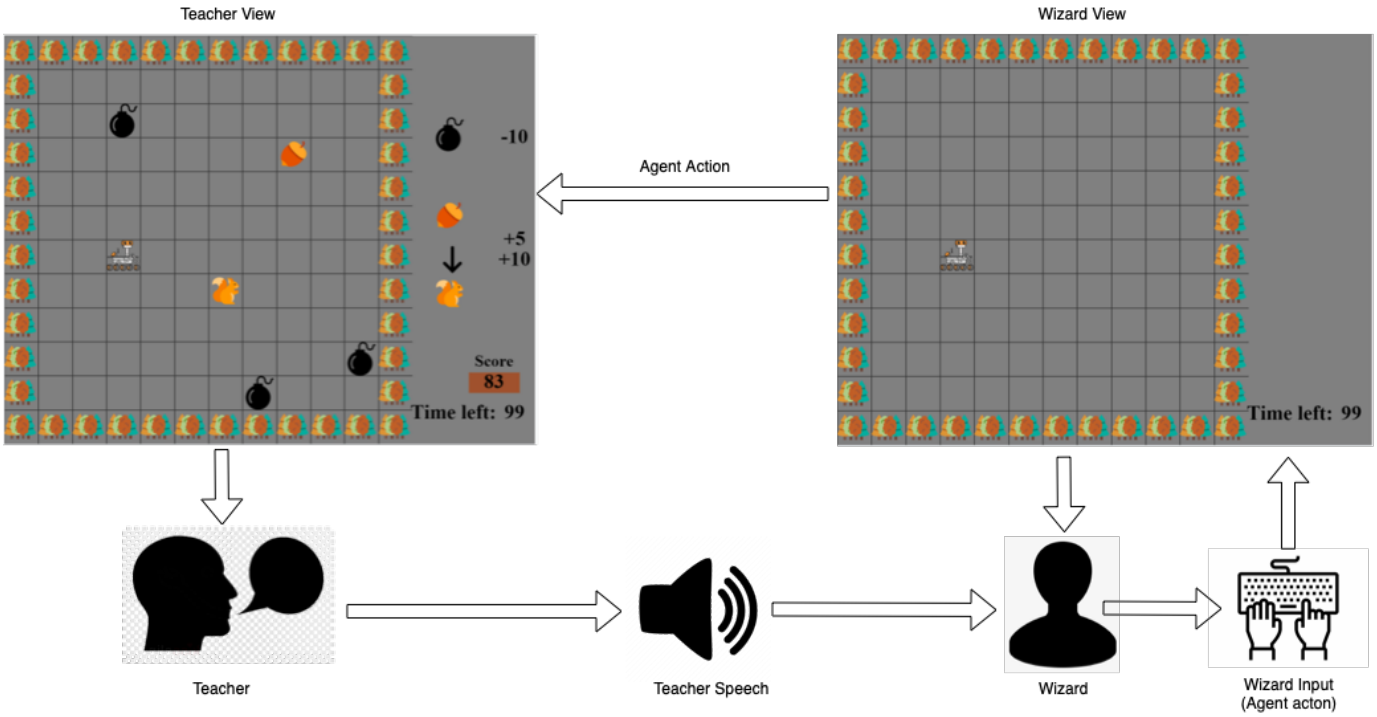

Fig. 1. Mixed Participant Wizard of Oz setup.

C. Procedure

Pairs of participants were appointed simultaneously for the study and welcomed by the examiner separately in either the teacher or the wizard role. All participants were presented with a relevant consent form prior to the session, with options for the teacher to choose how their data is shared according to privacy preferences. The study was approved by the university ethical committees of two of our affiliations.

1) Teacher: As a baseline recording of the teacher's voice prior to the start of the experimental part, the participant read a small paragraph given to them, which takes about 30 seconds to read. During the actual experiment, the teacher was only allowed to use the words "yes" and "no". To elicit richer prosodic variations, we instructed the teachers to speak as if they would with a 2 year old child. Before the experiment, the teacher was told that the agent would be listening to their voice, including "how" they spoke, and acting accordingly. In reality, the another participant in the role of Wizard was controlling the agent in response to the teachers' verbal feedback. The teacher was briefed about this deception after the experiment and was asked not to talk about this with other possible participants.

2) Wizard: The wizard was briefed on how to play the game. Their keyboard interactions were recorded alongside other details about the game such as: immediate rewards, timestamps, movements of the agent, and score. After the experiment, the wizard was asked not to talk about the experiment procedure with other potential participants.

D. Measures

Given the scope of the paper, we only focus on measures of the teacher's data. Analysis of the Wizard data is left for future work.

1) Prosodic features: We computed several acoustic features over the detected utterances to characterize prosody from human verbal feedback. We studied three different features to capture prosody and affect – energy, loudness, and pitch. These acoustic features have been shown to enhance semantic parsing [34], understand speech recognition failures in dialogue systems [35], and widely used for applications in human-robot interaction [6], [36] and speech recognition [37] communities. These features were extracted from the audio recording of the experiment sessions. First, transcriptions of the recordings were created using Google Cloud's Speech-To-Text [38]. The transcriptions were used to filter out silent parts and speech other than "yes" and "no" (which mostly consisted of non-verbal sounds). After correctly identifying where the relevant speech took place, prosody features were calculated. Librosa [39] was used for calculating the loudness and energy of the audio. Pitch (or frequency) was analyzed using a pre-trained model for a data-driven pitch tracking algorithm [40]. During the analysis, mean, max, and raw values were used as candidate measures.

2) MDP features: As is most reinforcement learning settings, we model the learning task as a Markov Decision Process (MDP). The Q-value $Q(s, a)$ representing the expected utility of action a at state s if the agent follows the optimal policy. From it, the advantage function is calculated as:

Teacher	Baseline energy	Baseline pitch
1	$\mu = 0.011$ ($\sigma = 0.036$)	$\mu = 131.02$ ($\sigma = 58.33$)
2	$\mu = 0.013$ ($\sigma = 0.040$)	$\mu = 181.97$ ($\sigma = 35.33$)
3	$\mu = 0.022$ ($\sigma = 0.058$)	$\mu = 120.65$ ($\sigma = 41.32$)
4	$\mu = 0.002$ ($\sigma = 0.007$)	$\mu = 198.55$ ($\sigma = 33.45$)

TABLE I

MEAN AND STANDARD DEVIATION VALUES FOR PROSODIC FEATURES OF BASELINE RECORDINGS.

Teacher	$\rho_{\text{en.}\times\text{adv.}}$	$\rho_{\text{pitch}\times\text{adv.}}$	$\rho_{\mu\text{en.}\times\text{adv.}}$	$\rho_{\mu(\text{pitch})\times\text{adv.}}$
1	0.34^{**}	0.36^{**}	$\mathbf{0.54}^{**}$	$\mathbf{0.43}^{*}$
2	0.30^{**}	0.27^{**}	0.34^{**}	0.34^{**}
3	$\mathbf{0.47}^{**}$	0.36^{**}	0.23^{*}	0.30^{**}
4	$\mathbf{0.44}^{**}$	0.29^{**}	0.30^{*}	0.27^{**}

TABLE II

MAIN CORRELATION RESULTS BETWEEN PROSODIC FEATURES (RAW AND MEAN ENERGY AND PITCH) AND ADVANTAGE VALUE PER SESSION, WITH STRONG CORRELATIONS SHOWN IN BOLD. *: $p \leq 0.05$; **: $p < 0.001$

$A(s,a) = Q(s,a) - V(s)$. Recent work by Cui et al. [25] states that advantage might be a better task statistic to consider than the reward for analyzing social signals interpreted as evaluative feedback. Our hypothesis is that the advantage function, as a measure of relative performance of a given action at a given state, given the overall optimal policy, would significantly correlate with our prosodic features, with potential differences across different teachers based on their expressivity levels.

E. Data analysis

A total of 25 minutes and 7 seconds of audio was recorded from 4 sessions. These recordings include the baseline teacher audio and the audio recorded during the experiment. We conducted a correlation analysis on the advantage values and prosodic features that were extracted from the experiment sessions. Advantage values were calculated using policy iteration with a 0.98 discount factor. To create a semantically consistent signal, prosodic features from "no" parts were multiplied by -1 to indicate negative feedback. Spearman correlation was used to analyze the correlation between advantage values and prosodic features, as the relationship between the two was found to be non-linear.

V. RESULTS AND DISCUSSION

We briefly report on preliminary correlation results and some qualitative observations to guide our follow-up studies.

A. Correlation analysis

Our analysis for the four teachers with different baseline prosody (see Table I) is summarized in Table II. It shows four strong, seven moderate and five low correlations (as interpreted by standard statistics guidelines [41]), all statistically significant. The results also suggest some level of variability across different teacher in terms of which prosodic signal was used to modulate verbal feedback. This will be a subject of further investigation in a follow up study. Altogether, these results give us a good basis to believe we can potentially build a useful predictive model relating prosodic to MDP features with a larger sample size and a perhaps a richer feedback vocabulary.

B. Qualitative observations and lessons learned

Based on participants answers on open ended questions about their teaching experience, it became apparent that a transparency mechanism is needed on the agent side (e.g., backchanneling everytime the agent gets feedback) to communicate responsiveness and encourage the teacher to keep going. Another expected comment was that the teachers would have liked to tell the robot "what" to do by giving feedback as action advice. While for the sake of this work, we intentionally only focused on one type of feedback to minimize sources of noise, in the future we would like a prosody-sensitive algorithm to include multiple types of feedback, or even free-form speech to maximize usability and enhance the teaching experience.

VI. CONCLUSION AND NEXT STEPS

We introduced a mixed-participant Wizard-of-Oz paradigm, which we believe to be a promising approach for intRL researchers to study human teacher's social signals in a way that is scalable, flexible, and allows for real-time interaction. Furthermore, our pilot study on shows promising results in relating prosodic information such as energy and pitch in relation to the task at hand, namely the advantage function. Our goal is for an agent to be able use prosody as a teaching signal to improve its policy performance, which we wish to investigate in a future study. We are also interested in taking this work a step further with a larger dataset to build a generic computational model of prosody in relation to a learning task defined through an MDP. Further investigations with a larger participant pool, and a richer set of prosodic and MDP features may also allow us to build predictive models that account for the variation among people based on baseline speech data as well as, potentially, demographic information.

REFERENCES

[1] W. B. Knox and P. Stone, "Interactively shaping agents via human reinforcement: The tamer framework," in *Proceedings of the fifth international conference on Knowledge capture*. ACM, 2009, pp. 9–16.

[2] S. Griffith, K. Subramanian, J. Scholz, C. L. Isbell, and A. L. Thomaz, "Policy shaping: Integrating human feedback with reinforcement learning," *Advances in neural information processing systems*, vol. 26, 2013.

[3] T. Cederborg, I. Grover, C. L. Isbell, and A. L. Thomaz, "Policy shaping with human teachers," in *Twenty-Fourth International Joint Conference on Artificial Intelligence*, 2015.

[4] A. C. Tenorio-Gonzalez, E. F. Morales, and L. Villaseñor-Pineda, "Dynamic reward shaping: training a robot by voice," in *Ibero-American conference on artificial intelligence*. Springer, 2010, pp. 483–492.

[5] E. S. Kim and B. Scassellati, "Learning to refine behavior using prosodic feedback," in *2007 IEEE 6th International Conference on Development and Learning*. IEEE, 2007, pp. 205–210.

[6] E. S. Kim, D. Leyzberg, K. M. Tsui, and B. Scassellati, "How people talk when teaching a robot," in *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*, 2009, pp. 23–30.

[7] S. Krening, "Newtonian action advice: Integrating human verbal instruction with reinforcement learning," 2018.

[8] S. Krening, B. Harrison, K. M. Feigh, C. L. Isbell, M. Riedl, and A. Thomaz, "Learning from explanations using sentiment and advice in rl," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 9, no. 1, pp. 44–55, 2016.

[9] R. Zhang, F. Torabi, L. Guan, D. H. Ballard, and P. Stone, "Leveraging human guidance for deep reinforcement learning tasks," *arXiv preprint arXiv:1909.09906*, 2019.

[10] G. A. Rummery and M. Niranjan, *On-line Q-learning using connectionist systems*. University of Cambridge, Department of Engineering Cambridge, UK, 1994, vol. 37.

[11] S. P. Singh and R. S. Sutton, "Reinforcement learning with replacing eligibility traces," *Machine learning*, vol. 22, no. 1-3, pp. 123–158, 1996.

[12] R. S. Sutton, A. G. Barto *et al.*, *Introduction to reinforcement learning*. MIT press Cambridge, 1998, vol. 135.

[13] M. N. Nicolescu and M. J. Mataric, "Natural methods for robot task learning: Instructive demonstrations, generalization and practice," in *Proceedings of the second international joint conference on Autonomous agents and multiagent systems*, 2003, pp. 241–248.

[14] M. Pardowitz, S. Knoop, R. Dillmann, and R. D. Zollner, "Incremental learning of tasks from user demonstrations, past experiences, and vocal comments," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 37, no. 2, pp. 322–332, 2007.

[15] A. Saran, K. Desai, M. L. Chang, R. Lioutikov, A. Thomaz, and S. Niekum, "Understanding acoustic patterns of human teachers demonstrating manipulation tasks to robots," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022.

[16] A. Lockerd and C. Breazeal, "Tutelage and socially guided robot learning," in *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, vol. 4. IEEE, 2004, pp. 3475–3480.

[17] J. Lin, Z. Ma, R. Gomez, K. Nakamura, B. He, and G. Li, "A review on interactive reinforcement learning from human social feedback," *IEEE Access*, vol. 8, pp. 120 757–120 765, 2020.

[18] T. A. K. Faulkner, E. S. Short, and A. L. Thomaz, "Interactive reinforcement learning with inaccurate feedback," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 7498–7504.

[19] R. Zhang, F. Torabi, G. Warnell, and P. Stone, "Recent advances in leveraging human guidance for sequential decision-making tasks," *Autonomous Agents and Multi-Agent Systems*, vol. 35, no. 2, pp. 1–39, 2021.

[20] G. Li, R. Gomez, K. Nakamura, and B. He, "Human-centered reinforcement learning: A survey," *IEEE Transactions on Human-Machine Systems*, vol. 49, no. 4, pp. 337–349, 2019.

[21] A. Saran, S. Majumdar, E. S. Short, A. Thomaz, and S. Niekum, "Human gaze following for human-robot interaction," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 8615–8621.

[22] A. Saran, R. Zhang, E. S. Short, and S. Niekum, "Efficiently guiding imitation learning agents with human gaze," *International Conference on Autonomous Agents and Multiagent Systems*, 2021.

[23] A. Saran, E. S. Short, A. Thomaz, and S. Niekum, "Understanding teacher gaze patterns for robot learning," in *Conference on Robot Learning*. PMLR, 2020, pp. 1247–1258.

[24] R. Zhang, A. Saran, B. Liu, Y. Zhu, S. Guo, S. Niekum, D. Ballard, and M. Hayhoe, "Human gaze assisted artificial intelligence: A review," in *IJCAI: Proceedings of the Conference*, vol. 2020. NIH Public Access, 2020, p. 4951.

[25] Y. Cui, Q. Zhang, A. Allievi, P. Stone, S. Niekum, and W. B. Knox, "The empathic framework for task learning from implicit human feedback," *arXiv preprint arXiv:2009.13649*, 2020.

[26] J. Lin, Q. Zhang, R. Gomez, K. Nakamura, B. He, and G. Li, "Human social feedback for efficient interactive reinforcement agent learning," in *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2020, pp. 706–712.

[27] F. G. Visi and A. Tanaka, "Towards assisted interactive machine learning: exploring gesture-sound mappings using reinforcement learning," in *ICLI 2020—the fifth international conference on live interfaces*, 2020, pp. 9–11.

[28] F. Cruz, G. I. Parisi, and S. Wermter, "Multi-modal feedback for affordance-driven interactive reinforcement learning," in *2018 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2018, pp. 1–8.

[29] Z.-Y. Chen, Y.-J. Li, M. Wang, F. Steinicke, and Q. Zhao, "A reinforcement learning approach to redirected walking with passive haptic feedback," in *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 2021, pp. 184–192.

[30] K. Zhang, M. Sharma, M. Veloso, and O. Kroemer, "Leveraging multimodal haptic sensory data for robust cutting," in *2019 IEEE-RAS 19th International Conference on Humanoid Robots (Humanoids)*. IEEE, 2019, pp. 409–416.

[31] D. Gandhi, A. Gupta, and L. Pinto, "Swoosh! rattle! thump!-actions that sound," 2019.

[32] V. Dean, S. Tulsiani, and A. Gupta, "See, hear, explore: Curiosity via audio-visual association," *arXiv preprint arXiv:2007.03669*, 2020.

[33] Y. Aytar, T. Pfaff, D. Budden, T. L. Paine, Z. Wang, and N. de Freitas, "Playing hard exploration games by watching youtube," *arXiv preprint arXiv:1805.11592*, 2018.

[34] T. Tran, S. Toshniwal, M. Bansal, K. Gimpel, K. Livescu, and M. Ostendorf, "Parsing speech: a neural approach to integrating lexical and acoustic-prosodic information," *arXiv preprint arXiv:1704.07287*, 2017.

[35] J. Hirschberg, D. Litman, and M. Swerts, "Prosodic and other cues to speech recognition failures," *Speech communication*, vol. 43, no. 1–2, pp. 155–175, 2004.

[36] E. S. Short, M. L. Chang, and A. Thomaz, "Detecting contingency for hri in open-world environments," in *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, 2018, pp. 425–433.

[37] D. Yu and L. Deng, *AUTOMATIC SPEECH RECOGNITION*. Springer, 2016.

[38] Google, "Google Cloud Speech-to-Tex," https://cloud.google.com/speech-to-text, 2021.

[39] B. McFee, A. Metsai, M. McVicar, S. Balke, C. Thomé, C. Raffel, F. Zalkow, A. Malek, Dana, K. Lee, O. Nieto, D. Ellis, J. Mason, E. Battenberg, S. Seyfarth, R. Yamamoto, viktorandreevichmorozov, K. Choi, J. Moore, R. Bittner, S. Hidaka, Z. Wei, nullmightybofo, A. Weiss, D. Hereñú, F.-R. Stöter, L. Nickel, P. Friesch, M. Vollrath, and T. Kim, "librosa/librosa: 0.9.2," Jun. 2022. [Online]. Available: https://doi.org/10.5281/zenodo.6759664

[40] J. W. Kim, J. Salamon, P. Li, and J. P. Bello, "Crepe: A convolutional representation for pitch estimation," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 161–165.

[41] C. P. Dancey and J. Reidy, *Statistics without maths for psychology*. Pearson education, 2007.