ZIBELINE INTERNATIONAL™ PUBLISHING

ISSN: 2773-6202 (Online)
CODEN: JTIOCC

Journal of Technology & Innovation (JTIN)

DOI: http://doi.org/10.26480/jtin.01.2023.01.05

RESEARCH ARTICLE

DEEP REINFORCEMENT LEARNING WITH HIDDEN MARKOV MODEL FOR SPEECH RECOGNITION

Samson Isaac*, Khalid Haruna, Muhammad Aminu Ahmad, Rabi Mustapha

Department of Computer Science, Kaduna State University, Kaduna, Nigeria.
Corresponding Author Email: samson.isaac@kasu.edu.ng

This is an open access article distributed under the Creative Commons Attribution License CC BY 4.0, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ARTICLE DETAILS

Article History:

Received 28 November 2022
Revised 09 December 2022
Accepted 17 January 2023
Available online 19 January 2023

ABSTRACT

Nowadays, many applications uses speech recognition especially the field of computer science and electronics, Speech Recognition (SR) is the interpretation of words spoken into a text. It is also known as Speech-To-Text (STT) or Automatic-Speech-Recognition(ASR), or just Word-Recognition(WR). The Hidden-Markov-Model (HMM) is a type of Markov model, which means that the future state of the model depends on the current state, not on the entire history of the system and the goal of HMM is to learn a sequence of hidden states from a set of known states. The Long-Short-Time-Memory (LSTM) network is a type of Recurrent Neural Network (RNN) that can learn long-term dependencies between time steps of sequence data. The LSTM network is trained by the network in order to predict the values of subsequent time steps in a series-to-series regression. Deep Neural Network (DNN) models are better classifiers than Gaussian Mixture Models (GMMs), they can generalize much better with a smaller number of parameters over complex distributions. They model distributions of different classes jointly, called "distributed" learning, or, more properly "tied" learning. This work is aimed at developing a speech recognition model that will predict isolated speech of some selected fruits in Hausa, Igbo and Yoruba language by using the predicting power of Mel-Frequency-Cepstral-Coefficient (MFCC), LSTM and HMM algorithms. The findings of the study would improve the development of better automatic speech applications systems and would benefit the academic and research community in the field of Natural Language Processing.

KEYWORDS

Speech Recognition; Hidden Markov Model; Natural Language Processing; Deep Learning; Reinforcement Learning

1. INTRODUCTION

Artificial intelligence (AI) is becoming increasingly popular in many research fields, including computer engineering, healthcare, computer science, Natural Language Processing (NLP), and most importantly in signal processing. Audio processing covers several areas, such as processing speech, music, and ambient sound. Artificial intelligence (AI) techniques play a major role in all these areas moreso in the development of intelligent audio systems (Purwins et al., 2019). One of the main goals of artificial intelligence is to develop autonomous, voice-activated intelligent agents that listen to or interact with their environment to improve their behavior over time through trial and error. Designing such autonomous systems has long been a challenge, from robots that can adapt to changes in their environment to software agents that can communicate with humans using their language and natural environment. Reinforcement Learning (RL) provides a rule-based mathematical framework for such expert learning, although RL has had some success in the past (Sutton et al., 1998; Kohl and Stone, 2004; Ng et al., 2006). Previous methods were limited to small problems due to a lack of weight. In addition, RL also presents memory, computational complexity, and sampling issues when it comes to learning algorithms (Strehl et al., 2006).

Recently, Deep Learning (DL) models have emerged as new tools with robust features to evaluate features and provide instructions to solve these problems. The discovery of DL greatly increased productivity and had a great impact on many fields, from health to transportation and social sciences to biology. Deep Neural Networks (DNN), Convolutional Neural Networks (CNN) and Artificial Neural Networks (ANN) (Mohamed et al., 2009; Hinton et al., 2012; LeCun et al., 1989; Hochreiter and Schmidhuber, 1997). LSTM network are advanced techniques that enable many applications beyond natural audio signal processing. The application of DL algorithms to RL accelerated the development of DRL, this result to the field of Deep Reinforcement Learning (DRL). DRL uses improvements in DL to demonstrate the learning, power, and speed of RL algorithms. This allows RL to work across large scales and functional areas to solve previously problematic problems (Lange et al., 2012). DRL has been deployed in different applications such as general agents in complex travel environments, Natural Language Processing (NLP), land transportation, robots for policy management, computer vision and many others Audio signal processing is one of the specific area where the DRL is gaining more interest (Duan et al., 2016; Wang et al., 2016; Levin et al., 2016; Naeem et al., 2020; Zhao et al., 2019; Le et al., 2021; Arukumaran et al., 2017).

DRL has recently been used as a new tool to solve many problems and tasks related to Automatic Speech Recognition (ASR), System Response Sounds (SRS), Speech Recognition (SER), Audio Enhancement, Music Creation, and Voice Control. robot. While the output and applied of DRLs increased by 3-4 and 2-3 orders of magnitude between 2015 and 2022. There are many review articles on DRLs. A group researchers provide a brief overview of DRL, including important developments in DRL by illustrating new methods to use DNN to generate autonomous agents (Arul et al., 2017). Li also provides detailed information about DRL and includes several aspects of its application to compare its results and problems (Li, 2017). Other relevant studies include the implementation of DRL in communication networks, human-level agents and automated driving (Luong et al., 2019; Nguyen et al., 2020; Sallab et al., 2017). None of these

Quick Response Code	Access this article online	
	Website: www.jtin.com.my	**DOI:** 10.26480/jtin.01.2023.01.05

Cite The Article: Samson Isaac, Khalid Haruna, Muhammad Aminu Ahmad, Rabi Mustapha (2023). Deep Reinforcement Learning with Hidden Markov Model for Speech Recognition. *Journal of Technology & Innovation, 3(1): 01-05.*

articles discuss the use of DRLs in voice processing, as shown in Table 1 and Table 2. Natural Language Processing (NLP) is a tool to understand human languages that does not require machine learning. Some of the NLP used is Chabot's, voice assistants or procedural information (Lawrence and Giles, 2000). The research aims to propose a speech recognition system that recognizes different words by taking into account the miss rate and also enhance the prediction accuracy of the speakers.

2. Related Work

Automatic speech recognition (ASR) is the process of converting speech signals into good speech using algorithms. Today's ASR technology has reached a high level of performance that improves DL processes. However, the performance of the ASR system depends on the supervised training of deep models with a large amount of recorded data. The feature languages, the added cost of transcription makes ASR difficult to use for beginners. To expand the scope of ASR, several studies have tested DRL-based models for learning from feedback or environment. This form of learning aims to reduce the cost and time of transcription by giving positive or negative results instead of writing in full. For example, Kala and Shinozaki proposed an RL framework for ASR based on political theory that provides new insight to existing learning and classification methods (Kala and Shinozaki, 2018). This enables the ASR system to learn user feedback, which helps achieve better word recognition performance and a Word Error Rate (WER) compared to unsupervised classification.

In ASR, the flow-by-flow model is a great success. However, this model does not bring the language closer to the real-world during reference. A group researcher solved this problem by training the model step-by-step with the policy gradient algorithm (Tjandra et al., 2018). In contrast to the standard Maximum Probability Estimate (MPE), they use exponential techniques to describe all the scripts, increasing the negative distance directly as a gain. The results show a significant improvement with the RL-based target and the MPE target compared to the training model only with the MPE target. In another study, the authors found that using token-level games (indirect games provided after each step) sentences and level-level games (Tjandra et al., 2019). To address the sequence-by-sequence control of ASR models, a studied the REINFORCE algorithm, which rewards ASR for generating the most accurate sentences for matching and typing errors (Chung et al., 2020). Experimental study shows that the DRL-based method can effectively reduce the behavioral error rate from 10.4% to 8.7%.

According to the proposed to improve the ASR by using quantitative analysis based on the objective gradient function of the policy (Karita et al., 2018). This means that the expected WER of the prediction model is low. Therefore, the author confirms that the proposed method improves language recognition. A group researcher with maximum likelihood and gradient objectives, learning and adaptation through learning outcomes (Zhou et al., 2018). You can drive progress through training goals in real time and achieve a 4% to 13% improvement in end-to-end ASR. In, the author tries to solve the order-to-order problem by proposing a model based on the distribution of control and the gradient method of the policy that can directly measure the logarithmic probability of the correct answer Myth (Luo et al., 2017). They found a significant effect on low and medium ASRs. Some researchers provide a two-dimensional framework for understanding native language based on political context (Radzikowski et al., 2019). They studied warm and semi-warm start trends and found encouraging results for English by Japanese and Polish speakers.

The final ASR system will be large and difficult to achieve accuracy. The DRL method can be used to generate a model panel (Alamdari et al., 2018). ShrinkML is an RL-based development to reduce the shrink value for each class in an LSTM-based ASR model (Dudziak et al., 2019). They used RL to overcome the limitations of Single Value Decomposition (SVD) associated with the compression of the ASR method and data analysis. According to the results, the authors found that the RL model can eliminate the ASR process compared to the manual model. To save ASR time, evaluated the first business model with an RL-based scale (Rajapakshe et al., 2020). They found that DRL training resulted in faster adaptation and better recognition in less time than untrained areas. To overcome the convergence delay of the REINFORCE algorithm, The research mentioned above shows ways to improve the performance of ASR, which is to interact with the environment using DRL (Williams, 1992; Lawson et al., 2018). Despite these promising results, further research on the DRL algorithm is needed to develop an autonomous ASR system that can operate in real-world environments.

REINFORCE algorithm is very popular in ASR, so research is needed to investigate other DRL algorithms and show their suitability for ASR. Various DRL methods for speech have been investigated in many studies. According to children learn language according to reinforcement principles by associating words with meaning (Skinner et al., 1957). Based on their research, the authors can show that language acquisition is a group based on the basic interpretation of knowledge through environmental knowledge, creating knowledge and knowledge through experience. A group of researchers developed a parallel language learning mechanism in a maze world (Yu et al., 2018). Get the teacher's advice from the questions that you answer the management in words. Many other studies also study RL language learning methods (Sinha et al., 2019; Isaac, 2016; Hermann et al., 2017; Hill et al., 2018; Isaac et al., 2021).

3. Result and Discussion

The audio recordings of various pronunciations of fruits spoken in Nigeria language using a Microphone. The system developed for isolated word speech and was trained and tested for isolated words representing the Hausa, Igbo and Yoruba spoken words. Thus, the research work focuses on developing a model of the system to show case the real concept of the speech recognition in Nigeria Local dialect with more consideration of Hausa, Igbo and Yoruba using HMM approach. The Audio data will be converted from time domain to frequency domain represents in computer readable format, such as Waveform-Audio-File (WAV), MPEG-4-Audio-Layer (MP4) and Windows-Media-Audio (WMA)

The design method of the proposed system is shown in figure 1 below:

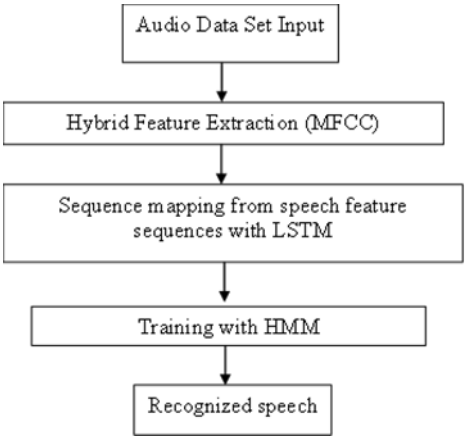

Figure 1: Designs Method

The system receives the isolated speech data set as input for feature extraction. The MFCC technique took advantage of the behavior of a speech signal. This method of determining a vector or value that can be used as an identity is called feature extraction. Because of its signal quality, MFCC is the most technology in many audio transmission applications. The feature is the coefficient of Cepstral, which is used when considering the human hearing system. MFCC functions is based on the different frequencies that can be received by the human ear to represent a human-like sound signal.

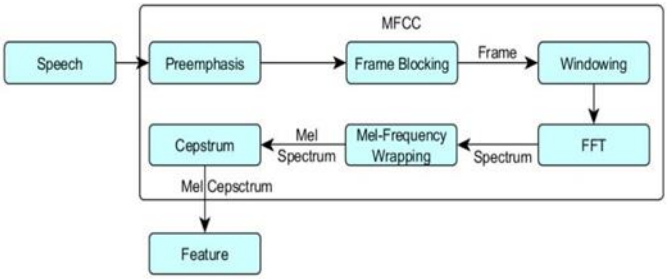

Figure 2: MFCC Feature Extraction Phase

LSTM is also a type of Recurrent Neural Network (RNN) that can learn long-term dependence between time series data. A series-to-series LSTM network is trained by the network to predict the values of the next series individually. LSTM works by accepting data input into the LSTM network as running data because it learns to predict future values at each step of the process. A Hidden Markov Model (HMM) is a statistical model that is assumed to be a Markov process with unknown parameters, and the goal is to determine the hidden parameters from the known data (Geiger et al., 2016). The known probabilities are combined with the HMM parameters in the HMM algorithm to create the Viterbi method for the best word order. Although, the state of HMM is not directly known, the state affects the variables. Each state has a probability distribution on the output signals.

Figure 3: The Working of the Proposed System

Table 1: Comparison of Speech Recognition Systems with Word Error Rate (WER)

Speech Recognition System	Dataset	WER (%)
HMM-GMM Vesely et al, (2013)	Switchboard	21.2
	CallHome	36.4
HMM-GMM Samr & Nizar, (2021)	TIMIT	N/A
MAP-GD-HMM Samr & Nizar, (2021)	TIMIT	N/A
DNN-HMM Vesely et al, (2013)	Switchboard	14.2
	CallHome	25.7
LSTM-HMM Saon et al, (2019)	Switchboard	7.2
	CallHome	12.7
TDNN-HMM Povey et al, (2016)	Switchboard	9.2
	CallHome	17.3
LSTM-CTC (Bigram Language Model) Graves & Jaitly, (2014)	Wall Street Journal	13.5
LSTM-CTC (No linguistic information) Graves & Jaitly, (2014)	Wall Street Journal	27.3
LSTM-CTC (Trigram language model) Graves & Jaitly, (2014)	Wall Street Journal	8.2
LSTM-CTC Graves & Jaitly, (2014)	Wall Street Journal	8.2
CNN-LSTM-CTC Zhang et al, (2017)	Wall Street Journal	10.5
LSTM-CTC Chiu et al, (2018)	Medical Dataset	20.1
Attention RNN-based Chan et al, (2016)	Wall Street Journal	10.3
Attention RNN-based Edward et al, (2017)	Medical Dataset	15.4
Attention RNN-based Chiu et al, (2018)	Medical Dataset	18.3
Transfomer Network Dong et al, (2018	Wall Street Journal	10.9
Transfomer Network Shigeki et al, (2019)	Wall Street Journal	4.5
Unimodal Model (Attention RNN-based) Srinivasan et al, (2020)	Flickr8K	13.7
Multimodal Srinivasan et al, (2020)	Flickr8K	13.4
Ours (LSTM-HMM)	Common Voice Hausa, Igbo, Yoruba	N/A

Table 2: Comparison of Speech Recognition Systems with Accuracy

Speech Recognition System	Dataset	Accuracy (%)
HMM-GMM Vesely et al, (2013)	Switchboard	N/A
	CallHome	
HMM-GMM Samr & Nizar, (2021)	TIMIT	50
MAP-GD-HMM Samr & Nizar, (2021)	TIMIT	93.33
DNN-HMM Vesely et al, (2013)	Switchboard	N/A
	CallHome	
LSTM-HMM Saon et al, (2019)	Switchboard	N/A
	CallHome	
TDNN-HMM Povey et al, (2016)	Switchboard	N/A
	CallHome	
LSTM-CTC (Bigram Language Model) Graves & Jaitly, (2014)	Wall Street Journal	N/A
LSTM-CTC (No linguistic information) Graves & Jaitly, (2014)	Wall Street Journal	N/A
LSTM-CTC (Trigram language model) Graves & Jaitly, (2014)	Wall Street Journal	N/A
LSTM-CTC Graves & Jaitly, (2014)	Wall Street Journal	N/A
CNN-LSTM-CTC Zhang et al, (2017)	Wall Street Journal	N/A
LSTM-CTC Chiu et al, (2018)	Medical Dataset	N/A
Attention RNN-based Chan et al, (2016)	Wall Street Journal	N/A
Attention RNN-based Edward et al, (2017)	Medical Dataset	N/A
Attention RNN-based Chiu et al, (2018)	Medical Dataset	N/A
Transfomer Network Dong et al, (2018	Wall Street Journal	N/A
Transfomer Network Shigeki et al, (2019)	Wall Street Journal	N/A
Unimodal Model (Attention RNN-based) Srinivasan et al, (2020)	Flickr8K	N/A
Multimodal Srinivasan et al, (2020)	Flickr8K	N/A
Ours (LSTM-HMM)	Common Voice Hausa, Igbo, Yoruba	96.62

4. Conclusion

When working with sequences of words, which is referred to as Natural Language Processing, LSTMs have had the best success. According to LSTM have a track record of carrying information while managing vanishing/exploding gradients. The findings of the study has improve the development of better automatic speech applications systems and would benefit the academic community in the field of Natural Language Processing. We develop a Speech Recognition system that has a 96.62 accuracy, low miss rate with high prediction accuracy for a vocabulary of some spoken words and various fruits classes.

References

Alamdari, N., Lobarinas, E., Kehtarnavaz, N., 2020. Personalization of hearing aid compression by human- in-the-loop deep reinforcement learning. IEEE Access 8, Pp. 203503–203515. https://doi.org/10.1109/ ACCESS.2020.3035728

Arulkumaran, K., Deisenroth, M.P., Brundage, M., and Bharath, A.A., 2017. Deep reinforcement learning: A brief survey. IEEE Signal Processing Magazine, 34 (6).

Chan, W., Jaitly, N., Le, Q., and Vinyals, O., 2016. Listen, attend and spell: A neural network for large vocabulary conversational speech recognition. In 2016 IEEE international conference on acoustics, speech and signal processing (ICASSP), Pp. 4960-4964.

Chiu, C.C., Tripathi, A., and Chou, K., 2018. Chris Co. Navdeep Jaitly, Diana Jaunzeikare, Anjuli Kannan, Patrick Nguyen, Hasim Sak, Ananth Sankar, Justin Tansuwan, Nathan Wan, Yonghui Wu, and Xuedong Zhang, Pp. 2972-2976. Speech recognition for medical conversations. arXiv:1711.07274.

Chung, H., Jeon, H.B., and Park, J.G., 2020. Semi-supervised training for sequence-to-sequence speech recognition using reinforcement learning. In: 2020 International Joint Conference on Neural Networks (IJCNN), Pp. 1–6.

Dong, L., Xu, S., and Xu, B., 2018. Speech-transformer: a no-recurrence sequence-to-sequence model for speech recognition. In 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Pp. 5884-5888.

Duan, Y., Schulman, J., Chen, X., Bartlett, P.L., Sutskever, I., and Abbeel, P., 2016. Rl2: Fast reinforcement learning via slow reinforcement learning. arXiv preprint arXiv, Pp. 1611.02779

Edwards, E., Salloum, W., Finley, G.P., Fone, J., Cardiff, G., Miller, M., and Suendermann-Oeft, D., 2017. Medical speech recognition: reaching parity with humans. In International Conference on Speech and Computer, Pp. 512-524.

Geiger, J.T., Zhang, Z., Weninger, F., Schuller, B., and Rigoll, G., 2016. Robust Speech Recognition using Long Short-Term Memory Recurrent NeuralNetworks for Hybrid Acoustic Modelling. Interspeech, Pp. 1-5.

Godin,W., Keith, and Hansen, H.L., John, 2015. Physical task stress and speaker variability in voice quality. Eurasip Jourrnal on Audio, Speech and Music Processing-(Springer), Pp. 1–13

Graves, A., and Jaitly, N., 2014. Towards end-to-end speech recognition with recurrent neural networks. In International conference on machine learning, pp. 1764-1772. PMLR.

Hermann, K.M., Hill, F., Green, S., Wang, F., Faulkner, R., Soyer, H., Szepesvari, D., Czarnecki, W.M., Jaderberg, M., Teplyashin, D., 2017. Grounded language learning in a simulated 3D world. arXiv preprint arXiv:1706.06551

Hill, F., Hermann, K.M., Blunsom, P., and Clark, S., 2018. Understanding grounded language learning agents.

Hinton, G., Deng, L., Yu, D., Dahl, G.E., Mohamed, A.R., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T.N., 2012. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. IEEE Signal processing magazine 29 (6).

Hochreiter, S., and Schmidhuber, J., 1997. Long short-term memory. Neural computation, 9 (8).

Isaac, S., Ahmed, M.A., Wisdom, D.D., Arinze, U.C., 2021. Battery-life management with an efficient sleep-mode power saving scheme (BM-ESPSS) in IEEE 802.16e networks. International Journal of Mechatronics, Electrical and Computer Technology (IJMEC), 11 (39), Pp. 4899-4904.

Isaac, S., 2016. Comparative Analysis of IPV4 and IPV6. International Journal of Computer Science and Information Technologies, 7 (2), Pp. 675-678.

Kala, T., Shinozaki, T., 2018. Reinforcement learning of speech recognition system based on policy gradient and hypothesis selection. In: International Conference on Acoustics, Speech and Signal Processing (ICASSP)

Karita, S., Ogawa, A., Delcroix, M., and Nakatani, T., 2018. Sequence training of encoder-decoder model using policy gradient for end-to-end speech recognition. In: International Conference on Acoustics, Speech and Signal Processing (ICASSP)

Kohl, N., Stone, P., 2004. Policy gradient reinforcement learning for fast quadrupedal locomotion. In: IEEE International Conference on Robotics and Automation (ICRA), 3.

Lange, S., Riedmiller, M.A., and Voigtländer, A., 2012. Autonomous reinforcement learning on raw visual input data in a real world application. In: International Joint Conference on Neural Networks (IJCNN), Brisbane, Australia.

Lawrence, S., and Giles, C.L., 2000. Natural Language Grammatical Inference with Recurrent Neural Networks. IEEE Transaction on Knowledge and Data Engineering, 12 (1), Pp. 126-140.

Lawson, D., Chiu, C.C., Tucker, G., Raffel, C., Swersky, K., Jaitly, N., 2018. Learning hard alignments with vari- ational inference. In: International Conference on Acoustics, Speech and Signal Processing (ICASSP)

Le, N., Rathour, V.S., Yamazaki, K., Luu, K., and Savvides, M., 2021. Deep reinforcement learning in computer vision: a comprehensive survey. Artificial Intelligence Review, Pp. 1–87.

LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., and Jackel, L.D., 1989. Backpropagation applied to handwritten zip code recognition. Neural computation, 1 (4).

Levine, S., Finn, C., Darrell, T., and Abbeel, P., 2016. End-to-end training of deep visuomotor policies. The Journal of Machine Learning Research, 17, (1).

Li, Y., 2017. Deep reinforcement learning: An overview. arXiv preprint arXiv:1701.07274

Luo, Y., Chiu, C.C., Jaitly, N., and Sutskever, I., 2017. Learning online alignments with continuous rewards policy gradient. In: International Conference on Acoustics, Speech and Signal Processing (ICASSP)

Luong, N.C., Hoang, D.T., Gong, S., Niyato, D., Wang, P., Liang, Y.C., and Kim, D.I., 2019. Applications of deep rein- forcement learning in communications and networking: A survey. IEEE Communications Surveys & Tutorials, 21 (4).

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., Graves, A., Riedmiller, M., Fidje- land, A.K., Ostrovski, G., 2015. Human-level control through deep reinforcement learning. Nature, 518 (7540).

Mohamed, A., Dahl, G., and Hinton, G., 2009. Deep belief networks for phone recognition. In: NIPS.

Naeem, M., Rizvi, S.T.H., and Coronato, A., 2020. A gentle introduction to reinforcement learning and its applica- tion in different fields. IEEE Access, 8, Pp. 209320–209344.

Ng, A.Y., Coates, A., Diel, M., Ganapathi, V., Schulte, J., Tse, B., Berger, E., and Liang, E., 2006. Autonomous inverted helicopter flight via reinforcement learning. In: Experimental robotics IX. Springer

Nguyen, T.T., Nguyen, N.D., Nahavandi, S., 2020. Deep reinforcement learning for multiagent systems: A review of challenges, solutions, and applications. IEEE transactions on cybernetics.

Povey, D., Peddinti, V., Galvez, D., Ghahremani, P., Manohar, V., Na, X., and Khudanpur, S., 2016. Purely sequence-trained neural networks for ASR based on lattice-free MMI. In Interspeech, Pp. 2751-2755.

Purwins, H., Li, B., Virtanen, T., Schlüter, J., Chang, S.Y., and Sainath, T., 2019. Deep learning for audio signal pro- cessing. IEEE Journal of Selected Topics in Signal Processing, 13 (2).

Radzikowski, K., Nowak, R., Wang, L., and Yoshie, O., 2019. Dual supervised learning for non-native speech recognition.EURASIP Journal on Audio, Speech, and Music Processing (1)

Rajapakshe, T., Latif, S., Rana, R., Khalifa, S., and Schuller, B.W., 2020. Deep reinforcement learning with pre-training for time-efficient training of automatic speech recognition. arXiv preprint arXiv:2005.11172

Sallab, A.E., Abdou, M., Perot, E., and Yogamani, S., 2017. Deep reinforcement learning framework for autonomous driving. Electronic Imaging, (19).

Samr, A., and Nizar, B., 2021. Maximum a posteriori approximation of hidden markov models for proportional sequential data modeling with simultaneous feature selection. IEEE Transactions on Neural Networks and Learning Systems, Pp. 1-12.

Saon, G., Kurata, G., Sercu, T., Audhkhasi, K., Thomas, S., Dimitriadis, D., and Hall, P., 2017. English conversational telephone speech recognition by humans and machines. arXiv preprint arXiv:1703.02136.

Shigeki K., Nelson, S., Shinji, W., Marc, D., Atsunori, O., and Tomohiro, N., 2019. Improving transformer-based end-to-end speech recognition with connectionist temporal classification and language model integration. INTERSPEECH 2019.

Silver, D., Huang, A., Maddison, C.J., Guez, A., Sifre, L., Van Den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., 2016. Mastering the game of go with deep neural networks and tree search. Nature, 529 (7587).

Sinha, A., Akilesh, B., Sarkar, M., and Krishnamurthy, B., 2019. Attention based natural language grounding by navigating virtual environment. In: IEEE Winter Conference on Applications of Computer Vision (WACV)

Skinner, B.F., 1957. Verbal behavior. new york: appleton-century-crofts Srinivasan, T., Sanabria, R., Metze, F., and Elliott, D., 2020. Multimodal

speech recognition with unstructured audio masking. arXiv preprint arXiv:2010.08642.

Strehl, A.L., Li, L., Wiewiora, E., Langford, J., and Littman, M.L., 2006. Pac model-free reinforcement learning. In: International Conference on Machine Learning (ICML)

Sutton, R.S., Barto, A.G., 1998. Introduction to reinforcement learning, vol. 135. MIT press Cambridge

Tjandra, A., Sakti, S., and Nakamura, S., 2018. Sequence-to-sequence ASR optimization via reinforcement learn- ing. In: International Conference on Acoustics, Speech and Signal Processing (ICASSP)

Tjandra, A., Sakti, S., and Nakamura, S., 2019. End-to-end speech recognition sequence training with reinforce- ment learning. IEEE Access 7

Veselý, K., Ghoshal, A., Burget, L., and Povey, D., 2013. Sequence-discriminative training of deep neural networks. In Interspeech, Pp. 2345-2349.

Wang, Z., Schaul, T., Hessel, M., Hasselt, H., Lanctot, M., and Freitas, N., 2016. Dueling network architectures for deep reinforcement learning. In: International Conference on Machine Learning (ICML)

Williams, R.J., 1992. Simple statistical gradient-following algorithms for connectionist reinforcement learning. Machine learning, 8 (3-4).

Yu, H., Zhang, H., and Xu, W., 2018. Interactive grounded language acquisition and generalization in a 2D world. In: International Conference on Learning Representations

Zhang, Y., Chan, W., and Jaitly, N., 2017. Very deep convolutional networks for end-to-end speech recognition. In 2017 IEEE international conference on acoustics, speech and signal processing (ICASSP), Pp. 4845-4849.

Zhao, T., Xie, K., and Eskénazi, M., 2019. Rethinking action spaces for reinforcement learning in end-to-end dialog agents with latent variable models. In: J. Burstein, C. Doran, T. Solorio (eds.) Conference of the North American Chapter of the Association for Computational Linguistics: Human Lan- guage Technologies (NAACL-HLT)

Zhou, H., Huang, M., Zhang, T., Zhu, X., and Liu, B., 2018. Emotional chatting machine: Emotional conversation generation with internal and external memory. In: AAAI Conference on Artificial Intelligence