Entrainable Neural Conversation Model Based on Reinforcement Learning

SEIYA KAWANO[1], **MASAHIRO MIZUKAMI**[2], **KOICHIRO YOSHINO**[1], (Member, IEEE),
AND SATOSHI NAKAMURA[1], (Fellow, IEEE)

[1]Graduate School of Science and Technology, Nara Institute of Science and Technology, Nara 630-0192, Japan
[2]NTT Communication Science Laboratories, Kyoto 619-0237, Japan

Corresponding author: Seiya Kawano (kawano.seiya.kj0@is.naist.jp)

ABSTRACT The synchronization of words in conversation, called entrainment, is generally observed in human-human conversations. Entrainment has a high correlation with dialogue success, naturalness, and engagement. In this article, we define entrainment scores based on the word similarities in semantic space to evaluate the entrainment of system generation. We optimized a neural conversation model to the entrainment scores using reinforcement learning so that the system can control the degree of entrainment of the system response. Experimental results showed that the proposed entrainable neural conversation model generated comparable or more natural responses than conventional models and satisfactorily controlled the degree of entrainment of the generated responses.

INDEX TERMS Neural conversation model, conditional response generation, entrainment, dialogue evaluation, reinforcement learning.

I. INTRODUCTION

Entrainment is a well-known conversational phenomenon in which dialogue participants mutually synchronize with regards to various aspects: lexical choice [1], syntax [2], style [3], acoustic prosody [4], [5], turn-taking [6], [7], and dialogue acts [8]. Entrainment has a high correlation with dialogue success, naturalness, and engagement [9]–[11]. Some existing works evaluated the dialogue quality and the performance of dialogue systems through entrainment analysis [12]. Although phenomena related to entrainment suggest that the quality of human-human and human-machine dialogues can be improved, it remains challenging to build a dialogue system that can explicitly consider the entrainment phenomena in the framework of a neural conversation model, which has been actively studied in recent years [13], [14].

In this article, we incorporate entrainment phenomena into a neural conversation model for building a more natural and user-satisfied dialogue system. We construct a neural conversation model that can control the degree of entrainment of generated responses based on a framework of reinforcement learning (RL) [15]. We define the automatic entrainment

The associate editor coordinating the review of this manuscript and approving it for publication was You Yang.

scores based on the local interpersonal distance [11], which focuses on lexical entrainment. We use this score to optimize a neural conversation model by RL.

In Section III, we describe our task of entrainable conversation modeling (Section III-A), a conditional generation model based on conventional architecture (Section III-B), and our proposed model optimized to entrainment scores by RL (Section III-C2). In experiments, we performed a preliminary analysis using the defined entrainment scores to clarify the relationship between user assessment and entrainment phenomena in a chit-chat dialogue domain (Section IV). Experimental results showed that our entrainment scores correlated with human assessment in human-human and human-machine dialogues in the chit-chat domain (Section V). As a model evaluation, we conducted subjective and objective evaluations (Section VI). Our proposed model generated comparable or more natural responses compared with general neural conversation models, which optimized by word prediction based on cross-entropy loss, and controlled well the degree of entrainment of the generated responses (Section VII). We discuss the challenges for the advancement of entrained response generation in neural conversation models by analyzing our experimental results (Section VIII).

II. RELATED WORKS

Many studies have analyzed entrainment in dialogues and shown that we can observe the phenomena in dialogues from various aspects: lexical choice [1], syntax [2], style [3], acoustic prosody [4], turn-taking [6], and dialogue acts [8]. Furthermore, automatic entrainment scores have been proposed that focuses on these aspects. These entrainment scores were highly correlated with dialogue success, naturalness, and engagement [9]–[11].

Some studies used the knowledge obtained by analyzing the entrainment to the dialogue system. One work [6] predicted the user's turn-taking behavior by considering entrainment. Another work [16] modeled a dialogue strategy to intentionally increase the accuracy of the automatic speech recognition using entrainment, and another [17] unified these works. Although these studies were conducted on modeling and predicting the entrainment of the user's behaviors, it remains challenging problem to build a dialogue system that can make entrainment to users to improve the dialogue system's response quality. In other words, insufficient studies have positively affected users through entrainment by the system.

On the other hand, recent neural conversation models focus on the efficient use and encoding of dialogue history [14], [18]. However, they do not directly handle entrainment phenomena because they are achieved by minimizing the cross-entropy loss of word prediction in decoder networks. Both the model networks that consider the dialogue context and the objective function of the model itself must be improved to achieve entrainable response generation.

In this article, we introduce a reinforcement learning (RL) framework [15], [19] to optimize a neural conversation model for automatic entrainment scores. Entrainment scores are given as RL rewards that enable neural conversation models to generate appropriately entrained responses for their dialogue contexts. Existing studies have already described the correlation between human assessments and automatic entrainment scores. We performed a follow-up analysis using chit-chat dialogue corpora to confirm that we can use the scores as an objective function. By optimizing the model to maximizing these scores, we expect that our neural conversation model can generate more natural responses.

III. ENTRAINABLE NEURAL CONVERSATION MODEL

In this article, we focus on lexical entrainment, which is related to lexical choice in dialogues. We introduce an entrainment score based on the similarity in semantic spaces in word-distributed representation [11], [20] to capture local entrainment trends for each turn in the dialogue. We optimize the neural conversation model using RL to maximize the entrainment scores of the generated responses. We formulate the problem as conditional neural conversation modeling, which uses the degree of entrainment of the response as a condition, because generating a highly entrained response is not always appropriate even though the model has to control the degree of entrainment based on the dialogue contexts.
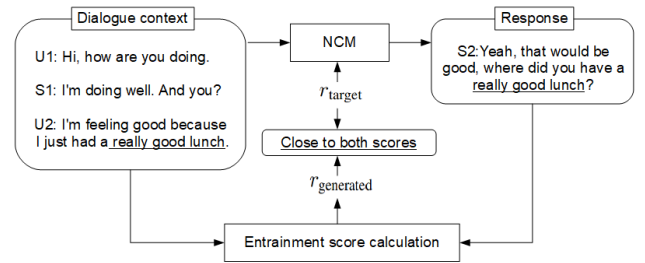

FIGURE 1. Task of entrainable conversation generation.

In this section, we first describe an overview of the response generation task tackled in this article (Section III-A). Then we describe the architecture of a conditional neural conversation model given the degree of entrainment (Section III-B). Finally, we describe a method that optimizes the conditional neural conversation model using RL to fit the given degree of entrainment (Section III-C).

A. TASK DEFINITION

We formally define the task of entrainable conversation modeling as a response generation task given a dialogue context and a degree of entrainment to the dialogue context. Define generated response word sequence $R = \{w_1, w_2, \cdots, w_T\}$, given dialogue context $H = \{H_1, H_2, \ldots, H_N\}$ and degree of entrainment of target response $r_{\text{target}} \in \mathbb{R}$. N is the dialogue length, and T is the number of words in an utterance.

In this setting, response R is required to satisfy not only the appropriateness to the dialogue context but also the degree of entrainment to the dialogue context (Fig. 1). In other words, the neural conversation model enforces entrainment degree $r_{\text{generated}} \in \mathbb{R}$ of the actual generated response to be closer to indicated entrainment degree r_{target}. This optimization is achieved by minimizing the relative error of both entrainment degrees:

$$\underset{r_{\text{generated}} \in \mathbb{R}}{\text{minimize}} \; \text{relative\_error}(r_{\text{target}}, r_{\text{generated}}). \quad (1)$$

As an approach to building such neural conversation models controllable by a given condition, such as the entrainment degree, vector concatenation is widely used between a word vector and the vectorized condition to feed a decoder input [21], [22]. Some other works proposed to extend models [23], [24] for conditional generation according to given emotion labels in the task of the emotional dialogue generation.

B. NEURAL CONVERSATION MODEL BASED ON ENTRAINMENT DEGREE

We introduce a conditional neural conversation model based on a hierarchical encoder-decoder model [14] with a context attention mechanism, which explicitly gives an embedded vector of entrainment degree to the decoder (Fig. 2). We apply the vector concatenation as a widely used method for conditioning the decoder.
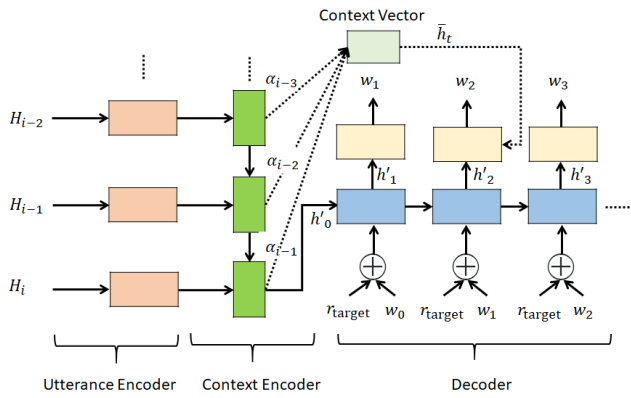
FIGURE 2. Neural conversation model with entrainment degree as a condition.

The encoder network has a hierarchical structure that consists of utterance and context encoders. The utterance encoder receives a word at each time step using forward RNNs to encode an utterance into a fixed-length vector:

$$h_{i,t} = \text{RNN}_{\text{utterance}}(h_{i,t-1}, \text{Embed}(w_{i,t})). \quad (2)$$

Here i is the number of turns in the dialogue context, and $h_{i,t}$ is the hidden vector obtained by inputting each word $w_{i,t}$ in utterance H_i. Each word $w_{i,t}$, which is encoded to a fixed-length vector using an embedding layer, is used as input.

In the context encoder, utterance vectors are input to encode the dialogue history:

$$c_i = \text{RNN}_{\text{context}}(c_{i-1}, h_{i,T}). \quad (3)$$

Here $h_{i,T}$ is a hidden vector obtained at the last step in the encoding for each utterance. Resultant vector c_i is fed into the decoder to generate a response sentence as initial hidden state h'_0. In the decoder, hidden state h' of the decoder and the output probability of word p_t are calculated:

$$h'_t = \text{RNN}_{\text{dec}}(h'_{t-1}, [\text{Embed}(w_{i,t-1}); \text{Linear}_{\text{ent}}(r_{\text{target}})]), \quad (4)$$
$$p_t = \text{softmax}(\text{Linear}_{\text{proj}}(h'_t)). \quad (5)$$

Here $\text{Linear}_{\text{proj}}$ is a projection layer, which maps h'_t to a vector of vocabulary size $|\mathcal{V}|$. $\text{Linear}_{\text{ent}}$ is a linear transformation layer that embeds target entrainment degree r_{target} into a fixed-length vector. $w_{i,t}$ is sampled from p_t and used as a part of the input for the next step. In this decoding architecture, we expect the decoder to generate a response with an appropriate degree of entrainment for the dialogue history by also inputting entrainment degree r_{target} in addition to already generated words. Note that we used teacher-forcing in the training process [13].

We also introduce a simple attention mechanism to the above decoder for efficiently handling the information from the encoded dialogue context. Specifically, let $c_{1:N}$ be a sequence of vectors obtained by the context encoder, and let

h'_t be the hidden states of the decoder in the t step. We compute the alignment weights based on general-attention [25] for each hidden state and obtain context vector $\bar{h}_t$:

$$\alpha_j = \frac{\exp(\text{score}(c_j, h'_t))}{\sum_{\tilde{j}=1}^{N} \exp(\text{score}(c_{\tilde{j}}, h'_t))}, \quad (6)$$

$$\bar{h}_t = \sum_{\tilde{j}=1}^{N} \alpha_{\tilde{j}} \cdot c_{\tilde{j}}. \quad (7)$$

The output words in step t are predicted using computed context vector $\bar{h}_t$:

$$\hat{h}_t = \tanh(\text{Linear}_{\text{attn}}([\bar{h}_t; h'_t])), \quad (8)$$
$$p_t = \text{softmax}(\text{Linear}_{\text{proj}}(\hat{h}_t)). \quad (9)$$

In general, training the neural conversation model is based on minimizing the cross-entropy:

$$L_{\text{CE}} = -\sum_{t=1}^{T} \log \frac{\exp(x_{t,e})}{\sum_{k}^{|\mathcal{V}|} \exp(x_{t,k})}. \quad (10)$$

Here $|\mathcal{V}|$ denotes the vocabulary size, $x_t \in \mathbb{R}^{|\mathcal{V}|}$ denotes the output of the projection layer in the decoding steps, and $x_{t,e} \in \mathbb{R}^{|\mathcal{V}|}$ denotes the e-th element that correspond to target word w_t.

However, perhaps models based on minimizing cross-entropy loss do not efficiently use the information in the dialogue context [26]. Furthermore, since cross-entropy loss is not designed to handle entrainment phenomena, we have to define a new objective function for building a neural conversation model that is optimized to entrainment scores.

C. MODEL OPTIMIZATION TO ENTRAINMENT DEGREE BASED ON REINFORCEMENT LEARNING

Our final goal is to build an entrainable neural conversational model based on the given entrainment degree. However, model optimization based on existing cross-entropy loss does not satisfactorily control the generation because the optimization is calculated word-by-word. In contrast, optimization based on reinforcement learning has the potential to train such a controllable response generation model [27]. Thus, we introduce the REINFORCE algorithm, which is based on reinforcement learning [15], [19].

In this section, we describe the objective function and its optimization method using the REINFORCE algorithm (Section III-C1). We introduce a reward function using an automatic entrainment score to optimize the model (Section III-C2) and scrutinize the training procedure for our neural conversation model based on reinforcement learning (Section III-C3).

1) REINFORCE ALGORITHM

The generation process in the neural conversation model is formalized as a Markov decision process (MDP) and optimized with reinforcement learning (RL) [28]. The problem of response generation in the neural conversation model is

generating response word sequence $R = \{w_1, w_2, \cdots, w_T\}$ that corresponds to given dialogue history H and target entrainment degree r_{target}. Formally the generation process is defined as choosing an action to generate word w_t given a state, already generated words $\{w_1, w_2, \cdots, w_{t-1}\}$, in time-step t [29]. Such a word selection process in the generation is defined as an action sequence, which is generated by an actual policy in MDP.

We define a reward function based on entrainment scores to encourage the model to generate entrained responses. The entrainment's evaluation score is fed as a reward to update the generator's policy in the RL. We use a policy gradient (REINFORCE algorithm) [15], [19] to train the policy. The objective function and its gradient are defined:

$$J_{\text{RL}}(\theta) = \sum_{w_{1:T}} G_\theta(w_t | w_{1:t-1}) \cdot Q^{G_\theta}(w_{1:t-1}, w_t), \quad (11)$$

$$\nabla J_{\text{RL}}(\theta) \simeq \frac{1}{T} \sum_{t=1}^{T} \sum_{w_t \in \mathcal{V}} Q^{G_\theta}(w_{1:t-1}, w_t)$$
$$\cdot \nabla_\theta G_\theta(w_t | w_{1:t-1}) \quad (12)$$
$$= \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{w_t \sim G_\theta} [Q^{G_\theta}(w_{1:t-1}, w_t)$$
$$\cdot \nabla_\theta \log p(w_t | w_{1:t-1})]. \quad (13)$$

Here θ is a parameter of the policy. $\mathcal{V}$ is a vocabulary, $w_{1:t-1}$ indicates the already generated word sequence (state in MDP), and $p(w_t | w_{1:t-1}) = G_\theta(w_t | w_{1:t-1})$ is the generative probability of word $w_t \in \mathcal{V}$ (action in MDP) in the decoder. $Q^{G_\theta}(w_{1:t-1}, w_t)$ is an action-value function that gives an expected future reward when the system generates word w_t given the state: already generated word sequence $w_{1:t-1}$.

The action-value function for each step is calculated using a Monte Carlo tree search (MCTS) [29], [30] under the current policy and its parameter θ:

$$Q^{G_\theta}(R_{1:t-1}, w_t)$$
$$= \begin{cases} \frac{1}{N} \sum_{n=1}^{N} r(H, R_{1:t}^n, R^{\text{ref}}, r_{\text{target}}) & \text{for } t < T, \\ r(H, R_{1:t}, R^{\text{ref}}, r_{\text{target}}) & \text{for } t = T. \end{cases} \quad (14)$$

Here $r(\cdot)$ is a reward function that evaluates the entrainment degree of response $R_{1:T} = \{w_1, w_2, \cdots, w_T\}$. $R_{1:t}^n$ is the generated response using a roll-out [29] from partial-generated response $R_{1:t}$ using parameter θ. R^{ref} is a reference response, and n is the number of roll-outs.[1] This reward function calculates the reward based on the relative error between a given entrainment degree and the entrainment degree of the actual generated response to allow it to control the entrainment degree of the generated response. Note that we can use an arbitrary score in this formulation. We use a reliable entrainment score based on the similarity of the semantic space of the words to feed the entrainment degree (score) as the reward.

2) REWARD FUNCTION FOR EVALUATING ENTRAINMENT DEGREE

We construct a reward function based on the idea of a local interpersonal distance (LID), which is a previously proposed turn-level entrainment score [11]. LID uses a predefined number of turns (context lengths) in response to the utterances of the primary speaker (anchor). The anchor utterance and response pair that has a minimum distance is chosen to calculate LID. This calculation is based on local entrainment, which is not always observed in the immediate response to the primary speaker's turn. It might be sustained and exhibited after a few turns [31]. In this article, unlike the LID's original definition [11], we calculate the similarity between the anchor utterance and each past contextual utterance by another speaker and choose an anchor and contextual utterance pair with minimum distance. However, note that there is no difference in the nature of both scores.

To calculate the distance between two utterances, we use Word Mover's Distance (WMD) [20], which is calculated from the distributed vector representations of words in a document. WMD targets both semantic and syntactic information to get a distance between text documents. WMD calculates the Earth Mover's Distance [32] between sets of word vectors that are contained in the target sentences (documents). This calculation is based on the minimum travel distance. Specifically, let $e_i \in \mathbb{R}^d$ represents i-th word, as defined by word-embedding $E \in \mathbb{R}^{d \times n}$ for vocabulary of n words. We also define a and b are n-dimensional normalized vectors, which consist of bag-of-words of two sentences. a_i indicates the count of the word i in the sentence.[2] The WMD introduces an transport matrix $T \in \mathbb{R}^{n \times n}$, such that $T_{i,j}$ indicates how much of a_i should be transported to b_j. Formally, the WMD learns T to minimize:

$$\text{WMD}(a, b) = \min_{T \geq 0} \sum_{i=1}^{n} \sum_{j=1}^{n} T_{ij} ||e_i - e_j||$$

$$\text{subject to} \sum_{j=1}^{n} T_{ij} = a_i \quad \forall i,$$

$$\text{and} \sum_{i=1}^{n} T_{ij} = b_j \quad \forall j. \quad (15)$$

To solve this minimization problem, we used the efficient implementations [20], [33], [34].[3]

We define target entrainment degree r_{target} based on the idea of LID:

$$\text{sim}(x, y) = e^{-\text{WMD}(\text{bow}(x), \text{bow}(y))^2}, \quad (16)$$

$$r_{\text{target}} = r_{\text{target}}(H, R^{\text{ref}}) = \max_{H_j^{\text{other}} \in H^{\text{other}} \subset H} \text{sim}(H_j^{\text{other}}, R^{\text{ref}}). \quad (17)$$

Here the sim$(\cdot)$ function normalizes the calculation results of WMD as the similarities between utterances (x and y). e

---

[1] We set the number of roll-outs to 5. However, since the computation cost of MCTS is high when training the large model, we can also adopt an approach for speeding up the training, such as REGS [28], instead of MCTS.

[2] Note that a_i is normalized over all words in a.

[3] https://github.com/RaRe-Technologies/gensim

is a natural logarithm, and bow(·) is a function to convert the given sentence to bag-of-words representation. R^{ref} is the reference response corresponding to dialogue history H, and $H^{\text{other}} \subset H$ is a set consisting of the most recent k utterances from the dialogue history H, excluding the primary speaker's utterances. r_{target} is assumed to be a similarity that takes values from 0 to 1.

Next we define entrainment degree $r_{\text{generated}}$ of the actual generated response:

$$u_{\text{target}} = u_{\text{target}}(H, R^{\text{ref}}) = \underset{H_j^{\text{other}} \in H^{\text{other}} \subset H}{\arg\max} \ \text{sim}(H_j^{\text{other}}, R^{\text{ref}}),$$ (18)

$$r_{\text{generated}} = r_{\text{generated}}(H, R^{\text{ref}}, R) = \text{sim}(R, u_{\text{target}}).$$ (19)

Here u_{target} is a target utterance to make entrainment by system generation, which has the maximum similarity of every pair formed by the anchor utterance of the reference and a context utterance. Thus, $r_{\text{generated}}$ is calculated as the similarity between the generated response and the target.

The reward given to the generated response is calculated from the relative error between r_{target} and $r_{\text{generated}}$:

$$r = r(H, R, R^{\text{ref}}, r_{\text{target}}) = 1 - \frac{|r_{\text{target}} - r_{\text{generated}}|}{\max(r_{\text{target}}, 1 - r_{\text{target}})}.$$ (20)

This reward function gives more reward when the relative error between the entrainment degree of the generated response and the indicated entrainment degree is small. In other words, it gives penalty if the generated utterance is over or under-entrained compared with the reference.

We used different functions for r_{taget} and $r_{\text{generated}}$ because using the same function will lead to learning a lazy policy that always refers to the previous utterance.

3) MODEL TRAINING BASED ON REINFORCE ALGORITHM

The training procedure of a neural conversation model with RL is shown in Algorithm 1.

Algorithm 1 Training Procedure

Require: generator policy G_θ; roll-out policy G_θ
1: Initialize G_θ with random weights θ
2: Pretrain G_θ to minimize L_{CE} ▷ (10)
3: **for** number of iterations **do**
4:    $G'_\theta \leftarrow G_\theta$
5:    **for** number of steps **do**
6:       sample $(H, R^{\text{ref}}, r_{\text{target}})$ from training data
7:       generate response R using G'_θ on H and r_{target}
8:       compute Q^{G_θ} for $(H, R, r_{\text{target}})$ using G'_θ ▷ (14)
9:       update G_θ based on $J_{\text{RL}}(\theta)$ ▷ (11)

First, we pre-train the neural conversation model by minimizing cross-entropy loss L_{CE}. Then we train it to maximize objective function $J_{\text{RL}}(\theta)$ using reinforcement learning and add L_{CE} to the loss to stabilize the training. This approach works as a teacher forcing and prevents the collapse of

TABLE 1. Number of dialogues/utterances in each corpus.

Corpus	Dialogues	Utterances
ConvAI2-wild-evaluation	2,483	41,415
NTT-chit-chat	3,483	56,566

policies due to the model's inability to access the reference response [28]. The policy used to calculate Q^{G_θ} is updated every 20 steps. We use the model with the highest reward sum for the generated response and the deterioration of perplexity within 1.0 points in the validation set for the evaluation.

IV. ENTRAINMENT ANALYSIS SETTING

LID, which is used as a reward in this article, probably correlates with human assessment (therapeutic outcomes and affective behaviors) in the dialogues of clinical psychology and psychotherapy [11]. On the other hand, no examination has focused on chit-chat dialogues, which are the main focus of this article. Therefore, we performed a preliminary analysis to clarify the relationship between user assessment and entrainment in chit-chat dialogues.

We used Spearman's rank correlation coefficient to evaluate the relationship between the Conversational Linguistic Distance (CLiD) [11], a dialogue session-level entrainment score calculated by the mean of the LIDs, and the user assessment assigned to each dialogue. CLiD is defined:

$$\text{CLiD}(\mathcal{D}) = \frac{\sum_{(H, R^{\text{ref}}) \in \mathcal{D}} r_{\text{target}}(H, R^{\text{ref}})}{|\mathcal{D}|}.$$ (21)

We applied (17) to each turn of the dialogues and averaged the results as the dialogue level entrainment scores. Here, $(H, R^{\text{ref}}) \in \mathcal{D}$ is a context and response pair for each turn in the dialogue. Note that the definition changes from the original CLiD to fit our problem.

For the entrainment analysis, we used the following chit-chat dialogue corpora:

- **ConvAI2-wild-evaluation**: Dialogues between a human and a system that participated in the ConvAI2 competition.[4] Each dialogue was evaluated by human participants on a five-point scale.
- **NTT-chit-chat**: Dialogues between human participants that covered as wide range of topics [35]. Participants in each dialogue evaluated it from the following three viewpoints on a seven-point Likert scale: 1) strongly disagree; 2) disagree; 3) Slightly disagree; 4) neither; 5) slightly agree; 6) agree; 7) strongly agree.
  - Q1: "I am satisfied with the current dialogue. I'd like to have such a dialogue again."
  - Q2: "I found myself interested in the topic of the current dialogue."
  - Q3: "In the current dialogue, I spoke positively on my own."

---

[4]http://convai.io/data

TABLE 2. Entrainment analysis results for ConvAI2-wild-evaluation.

Types	$k = 1$		$k = 2$	
	ρ	p-value	ρ	p-value
Human→System	0.19	6.25×10^{-22}	0.24	2.83×10^{-33}
System→Human	0.22	9.22×10^{-31}	0.23	5.51×10^{-31}

TABLE 3. Entrainment analysis results for NTT-chit-chat.

Question	$k = 1$		$k = 2$	
	ρ	p-value	ρ	p-value
Q1	0.13	2.45×10^{-15}	0.10	7.13×10^{-10}
Q2	0.11	6.30×10^{-12}	0.08	4.11×10^{-7}
Q3	0.09	4.19×10^{-8}	0.07	2.66×10^{-5}

V. ENTRAINMENT ANALYSIS RESULTS

We performed a correlation analysis of ConvAI2-wild-evaluation, as shown in Table 2. Here ρ is the correlation calculated by Spearman's rank correlation analysis, and the p-value is the probability of the null hypothesis. We calculated the CLiD for two types based on their attributes because the speaker has distinctly different attributes: "Human → System" shows the human responses to the system, and "System → Human" shows the system responses to the humans.

As shown in Table 2, we confirmed that CLiD has a positive correlation with human scores, regardless of any setting used to calculate it. This result indicates that the entrainment degree is critical for improving the quality of human-machine dialogues. Since many systems based on a neural network are not able to handle the dialogue context [26], this result might be deeply related not only to entrainment but also to whether the system can generate a context-relevant response. To compare cases using different values of k, we confirmed a stronger correlation in the case of $k = 2$. In fact, humans often respond with an awareness of both the previous utterances but also deeper utterances from the past in a dialogue history [31].

We also performed a correlation analysis between average assessments by two participants and CLiD in an NTT-chit-chat (Table 3).

Table 3 shows a moderately positive correlation between the user assessments corresponding to these questions and the CLiD. This result suggests that using entrainment in dialogues is an effective strategy to improve user satisfaction in chit-chat dialogues. As stronger correlation is observed in $k = 1$ than in $k = 2$. This is probably because NTT-chit-chat has multiple utterances per turn. In other words, when we apply CLiD to dialogues that contain a lot of information in one turn, it is difficult to find strong correlations between CLiD and human scores, because LIDs, which are CLiD components, will be biased by the number of words in the utterances.

These results indicate that CLiD, which is calculated by LID averages, is a useful and strict score to evaluate dialogues in the chit-chat domain, if their utterance length is limited.

TABLE 4. Number of dialogues/utterances in ConvAI2 dataset.

	Dialogues	Utterances
Train	17,876	262,862
Validation	498	7,798
Test	499	7,788

In other words, these results support our hypothesis: maximizing the LIDs in dialogues increases dialogue quality.

VI. EVALUATION SETTING FOR RESPONSE GENERATION

We confirmed that improving LID scores is important in human-machine dialogue setting as well, which is a basic hypothesis of our entrainable dialogue modeling. In this section, we describe the experimental settings to confirm the effect of our proposed entrainable neural conversation model.

A. DATASET

We used the dataset provided at the ConvAI2 competition, which was also used to train the system in the ConvAI2-wild-evaluation described in Section IV. This dataset was divided into train, validation, and test sets (Table 4). We divided the original development data into validation and test sets.[5]

B. COMPETING MODELS

We compared the following different types of neural conversation models in our evaluations:

- **ASEQ2SEQ**: a standard neural conversation model that encodes a previous utterance as a query for decoding a response with an attention mechanism (general-attention) [25].
- **HED**: a hierarchical encoder-decoder model [14] without an attention mechanism and conditioning to a decoder.
- **AHED**: a model that combines an attention mechanism with the HED model.
- **C-ASEQ2SEQ**: a model with conditioning based on ASEQ2SEQ. We gave the condition (degree of entrainment) as described in Section III-B.
- **C-HED**: a HED model with a conditioning mechanism. We gave the condition (degree of entrainment) as described in Section III-B.
- **C-AHED**: an AHED model with conditioning. We gave the condition (degree of entrainment) as described in Section III-B.

We trained these neural conversation models using conventional cross-entropy loss ($+\textbf{CE}$) and our proposed optimization based on reinforcement learning ($+\textbf{RL}$). We used the entrainment scores as a condition given to the decoder and explored the case with different $k \in \{1, 2\}$ for score calculation. When $k = 1$, the entrainment score is calculated using only the previous utterance; when $k = 2$, it is calculated using the two most recent utterances by a non-primary speaker. Since the dialogue is performed alternately by two

---

[5]Note that the original test set in ConvAI2 dataset are private.

speakers, the neural conversation model needs to handle at most four utterances in a dialogue history when $k = 2$.

C. HYPER-PARAMETER SETTINGS

We used the same hyper-parameter settings in these models. The vocabulary size was 15000, the word embedding size was 300, the entrainment embedding size was 50, the hidden-size was 500. We used a two-layer Gated Recurrent Unit (GRU) [36] as an RNN. In the training, we used a mini-batch size of 128, and an Adam optimizer [37] with a learning rate of 1e-4. For the WMD calculation, we used pre-trained, word-distributed vectors,[6] which normalized the norm to 1 for each. We set the maximum length of the dialogue history to 4.

D. AUTOMATIC EVALUATIONS

We automatically evaluated the generation results using references in the test set. We used a beam search (a beam width of 5) for generating examples to be evaluated. For automatic evaluation, we used the following five different metrics:

- **Perplexity** (PPL) is a general metric for evaluating a language model performance. The model Likelihoods of the reference responses are calculated. Note that the perplexity scores do not directly reflect the quality of generation; for example, dull responses also have good perplexity scores.
- **BLEU**, which is the most popular automatic evaluation metric of language generation tasks, calculates the similarity between references and generated responses [38] based on n-gram precision. We used BLEU2, which considers uni-grams and bi-grams because matches in higher-order n-grams are rarely observed response generation tasks.
- $\overline{\text{WMD}}$ is the average similarity between the references and the generated responses for each case in the test set. The similarity is calculated based on (16). The score is multiplied by 100 and displayed in a range of 0 to 100.
- $\bar{r}$ is an average reward calculated from (20) to each generation. When this score is high, the entrainment degree of the generated response shows a similar degree to the reference. In other words, it shows that the neural conversation model controlled the response content well based on the entrainment degree. We sorted the test sets by the entertainment scores of the references and divided them into four parts to calculate $\bar{r}$ for each ($\bar{r}^{0\sim25\%}$, $\bar{r}^{25\sim50\%}$, $\bar{r}^{50\sim75\%}$, $\bar{r}^{75\sim100\%}$). For example, $\bar{r}^{0\sim25\%}$ shows the average reward of examples that have less entrainment scores in the references. These scores are multiplied by 100 and displayed in a range of 0 to 100.
- **Entropy** (Ent) is a diversity metric [39] that reflects the evenness of the empirical n-gram distribution for the given responses: $Ent = \frac{1}{\sum_{w \in \mathcal{V}} C(w)} \sum_{w \in \mathcal{V}} C(w) \log$

[6]For English data: http://nlp.stanford.edu/projects/glove/ and for Japanese data: http://www.worksap.co.jp/nlp-activity/word-vector/

$\frac{C(w)}{\sum_{w' \in \mathcal{V}} C(w')}$, where, $\mathcal{V}$ is the set of all n-grams in the given responses, and $C(w)$ denotes the frequency of n-gram w. We set the uni-gram for evaluation.

E. HUMAN SUBJECTIVE EVALUATIONS

Automatic evaluation scores still have a problem since they do not have high correlation with human subjective evaluation results [40]. Thus, we also examined models with a human subjective evaluation to confirm the naturalness of the generated responses. In the evaluation of naturalness, we used a 3-point scale in accordance with an existing work [41]. 240 generated responses were randomly selected from the test set, and human annotators added their evaluation scores for each sample by looking at the dialogue contexts. Detailed descriptions follow.

- **+2**: This response is not only relevant and natural, but also informative and interesting.
- **+1**: This response can be used as a response to the context, although it is universal, like "Yes, I see," "Me too," or "I don't know."
- **0**: This response cannot be used as a response to this context. It is either semantically irrelevant or dis-fluent.

Three annotators evaluated each sample, and the final score was decided by a simple majority. If the evaluation was completely disagreed (0, +1, and +2), the example was scored as 1.

VII. EXPERIMENTAL RESULTS ON ENTRAINED RESPONSE GENERATION

A. AUTOMATIC EVALUATION RESULTS

We show the automatic evaluation results of response generation models in Table 5. Our proposed models using the target entrainment degree as a condition showed improvement on $\bar{r}$ from the baseline models and achieved comparable performance on other metrics. Our proposed models controlled the generation at a high level, based on the indicated entrainment degree. In particular, we confirmed a remarkable improvement in models that applied reinforcement learning (C-HED+RL and C-AHED+RL). C-AHED+RL showed the best performance in $k = 2$, indicating that the attention mechanism for context works well when the model uses longer contexts. However, we still have a problem with generation performance $\bar{r}^{75\sim100\%}$, which gives very high entrainment scores as a condition. In other words, generating highly entrained responses is more challenging. Furthermore, our proposed models based on reinforcement learning showed a consistent improvement of $\overline{\text{WMD}}$ and Ent, and BLEU was comparable to the baseline models.

Then we traced the changes in the generation performance when we gave different fixed examples of r_{target} as a condition for the generation models instead of the oracle. The results are shown in Fig. 3. Here the vertical dashed line is the median of oracle r_{target}. Our proposed models, which are optimized by reinforcement learning, showed consistent improvement compared with the other models. For $\bar{r}$,

TABLE 5. Automatic evaluation results for each neural conversation model.

Models ($k=1$)	PPL	BLEU2	$\overline{\text{WMD}}$	$\bar{r}^{0\sim25\%}$	$\bar{r}^{25\sim50\%}$	$\bar{r}^{50\sim75\%}$	$\bar{r}^{75\sim100\%}$	$\bar{r}$	Ent
ASEQ2SEQ+CE	40.80	**7.82**	38.77	91.22	88.24	82.80	62.35	81.15	5.03
HED+CE	37.53	6.62	36.40	90.72	89.21	83.79	63.06	81.69	5.21
AHED+CE	39.11	6.71	38.16	91.04	88.44	82.40	62.11	81.00	5.11
C-ASEQ2SEQ+CE	40.05	3.69	38.77	90.90	88.08	84.19	68.82	83.01	5.33
C-HED+CE	37.12	6.56	39.17	91.41	88.52	84.25	70.04	83.55	5.20
C-AHED+CE	39.91	6.64	38.84	91.58	89.25	85.00	70.02	83.96	5.11
C-ASEQ2SEQ+RL	41.41	7.28	39.59	91.60	89.36	86.12	72.65	84.93	5.59
C-HED+RL	37.54	6.69	**40.57**	91.72	90.51	**88.01**	**74.53**	**86.19**	**5.97**
C-AHED+RL	40.01	**7.42**	39.96	**91.84**	**90.53**	86.36	72.75	85.37	5.52
Models ($k=2$)	PPL	BLEU2	$\overline{\text{WMD}}$	$\bar{r}^{0\sim25\%}$	$\bar{r}^{25\sim50\%}$	$\bar{r}^{50\sim75\%}$	$\bar{r}^{75\sim100\%}$	$\bar{r}$	Ent
HED+CE	37.53	6.62	36.40	87.99	86.69	82.45	62.74	79.97	5.21
AHED+CE	39.11	6.71	38.16	88.11	86.21	81.14	61.86	79.33	5.11
C-HED+CE	37.01	7.28	39.89	90.79	88.84	84.77	70.15	83.64	5.17
C-AHED+CE	38.76	5.17	38.43	89.49	87.64	84.44	67.54	82.27	5.18
C-HED+RL	37.24	6.39	**40.67**	90.25	89.49	86.91	**74.18**	85.21	5.62
C-AHED+RL	40.79	**7.30**	40.48	**90.93**	**91.24**	**87.14**	72.63	**85.49**	**5.87**

(a) Performance comparison when $k=1$

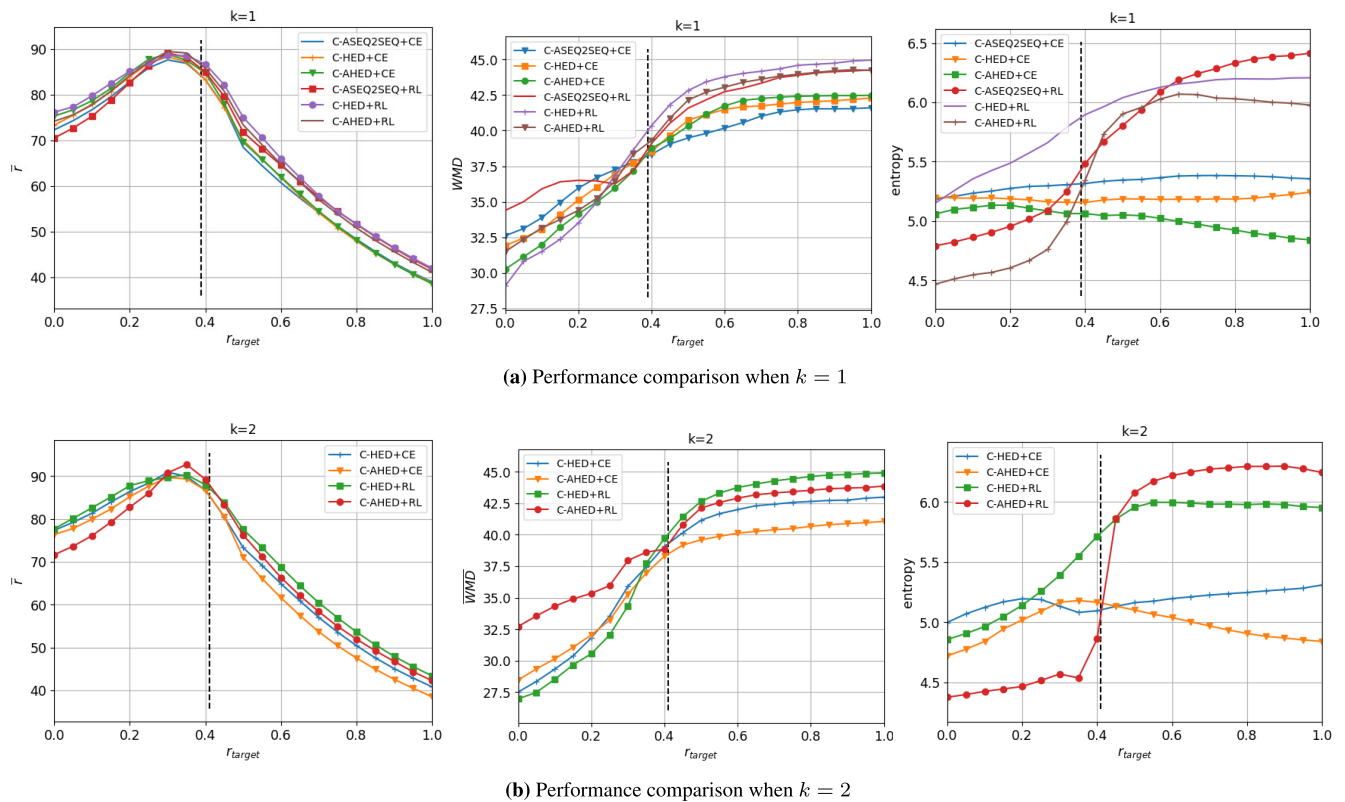**(b)** Performance comparison when $k=2$

FIGURE 3. Changes in generation performance when given a fixed r_{target}.

we confirmed the highest performance around the median of oracle r_{target}. Lower scores on high r_{target} were probably caused by a lack of training samples of high r_{target}. For $\overline{\text{WMD}}$ and Ent, we confirmed increasing trends in both scores when we give a high r_{target}. On the other hand, both scores are low in the range of low r_{target}. This result

was caused by dull responses, which have small diversity and little relevance to the references. In some cases, our models outperformed the results of giving the oracle conditions. This indicates that our models are robust even if the condition given to the model is different from the oracle.
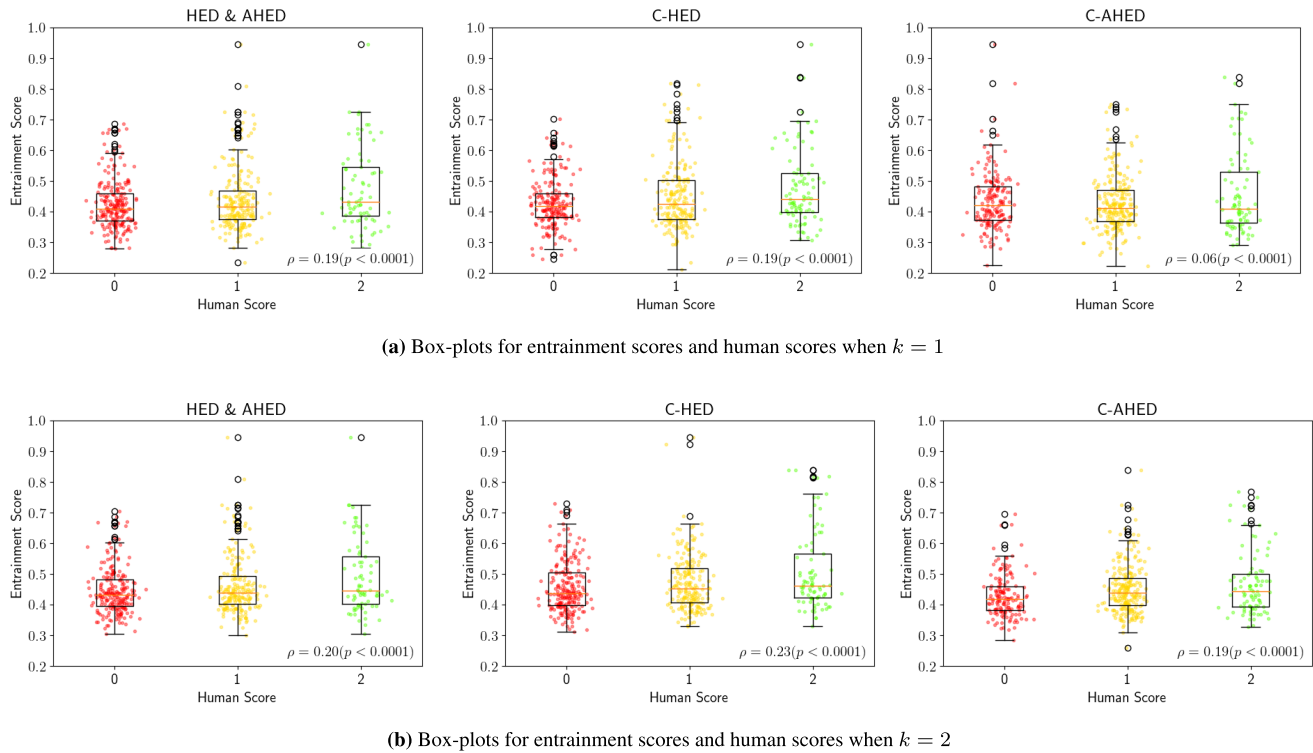
(a) Box-plots for entrainment scores and human scores when $k = 1$

(b) Box-plots for entrainment scores and human scores when $k = 2$

FIGURE 4. Relationships between human scores and entrainment scores.

B. HUMAN SUBJECTIVE EVALUATION RESULTS

Table 6 shows the human evaluation results for the naturalness of the generated responses in each model. We used the oracle entrainment degrees as the given conditions. Our proposed models, based on reinforcement learning, generated a more acceptable response to the dialogue context than the baseline models under the oracle condition. C-AHED+RL ($k = 2$) showed the highest performance. However, C-HED+RL ($k = 2$) did not improve the score compared with C-HED+RL ($k = 1$). This indicates that C-HED model, which has no attention mechanism, can not take enough advantage of reward signals that considering the more past context.

We also evaluated the relationship between human evaluation scores and entrainment scores based on LID.[7] Fig. 4 shows the box-plots for human evaluation scores and entrainment scores of three models (no-conditioned models and C-HEDs, C-AHEDs). Here the horizontal axis indicates the human evaluation scores and the vertical axis indicates the entrainment scores. Note that C-HEDs and C-AHEDs are including both results of cases in +CE and +RL. We calculated the polyserial correlation ρ [42] between human scores and entrainment scores instead of Spearman's rank correlation since there is only a 3-point scale for human scores. We identified significant positive correlations between human scores and entrainment scores for all of the

[7]The LID was calculated based on (17).

TABLE 6. Frequency of each subjective evaluation score. Weighted-Avg is a weighted average by frequency of scores.

Models	2	1	0	Weighted-Avg
HED+CE	44	85	111	0.72
AHED+CE	33	119	88	0.77
C-HED+CE ($k = 1$)	43	88	109	0.73
C-HED+CE ($k = 2$)	40	103	97	0.76
C-HED+RL ($k = 1$)	49	105	86	**0.85**
C-HED+RL ($k = 2$)	45	106	89	0.82
C-AHED+CE ($k = 1$)	45	105	90	0.81
C-AHED+CE ($k = 2$)	44	110	86	0.83
C-AHED+RL ($k = 1$)	38	123	79	0.83
C-AHED+RL ($k = 2$)	47	126	67	**0.92**

groups regardless of the k settings. This result is consistent with the results of the preliminary analysis in Section V. Note that we can not compare the magnitude of the correlations in each group. This is due to the correlations will be small for groups with a low frequency of score 0 since the nature of the evaluation score based on the 3-point scale.

Table 7 shows the generation examples of the compared models based on AHED in $k = 2$. Their naturalness seems at least comparable; even our proposed models generated more entrained responses. Note that it is difficult to find using the same words because our proposed method is based on WMD that calculates the semantic similarity in semantic spaces.

TABLE 7. Generated responses from models given a dialogue history and a target entrainment degree: A and B denote system and user. One utterance of each speaker is given as a history in the example. Generation results of each model are shown after the context. Before generation results, given target entrainment degrees and rewards bestowed on generated responses are indicated by brackets.

A	hi there ! how is your day ?
B: AHED+CE	i am doing well . how are you ?
B: C-AHED+CE	($r_{\text{target}} = 0.51$, $r = 0.79$) i am doing well . just got back from the gym . how are you ?.
B: C-AHED+RL	($r_{\text{target}} = 0.51$, $r = 0.98$) hello , i am good thanks . how are you ?
B: REFERENCE	($r_{\text{target}} = 0.51$, $r = 1.0$) i am doing great out here
A	have you thought about finding another job ?
B	i want to leave the job but my parents make me stay
A	maybe if you found a better one before leaving .
B	yeah but i'll move out my parents soon .
A	very cool ! i loved it when my kids left .
B	i'll love it when i leave then i can do as i like
A: AHED+CE	i am sorry to hear that . do you have any pets ?
A: C-AHED+CE	($r_{\text{target}} = 0.62$, $r = 0.72$) i am sorry to hear that . do you have any pets ?
A: C-AHED+RL	($r_{\text{target}} = 0.62$, $r = 0.83$) i hope so too ! what else do you like to do in your spare time ?
A: REFERENCE	($r_{\text{target}} = 0.62$, $r = 1.0$) it does not always work out that way
A	uh , i guess that is cool , so you like food ?
B	yup love it . i also survived cancer though .
A	bet that was pretty rough'
B	yes it was hard . i love to run . how about you ?
A	i spend a lot of time trying to cook , i'm not super good .
B	that is fine . i also like preparing for marathons .
A: AHED+CE	i love to cook , but i do not have a job , but i do not like it
A: C-AHED+CE	($r_{\text{target}} = 0.73$, $r = 0.45$) i am sorry to hear that . do you have any pets ?
A: C-AHED+RL	($r_{\text{target}} = 0.73$, $r = 0.55$) that is awesome , what kind of music do you like ?
A: REFERENCE	($r_{\text{target}} = 0.73$, $r = 1.0$) is that super hard work ?
A	i used to paint and would need to stop taking my meds to do a big job .
B	i understand . i like the city .
A	i am currently a cleaning lady and do alot of work in the city .
B	that sounds fun and interesting .
A	it is . what do you do ?
B	i work on construction right now .
A: AHED+CE	that sounds like a lot of fun . do you have any hobbies ?
A: C-AHED+CE	($r_{\text{target}} = 0.23$, $r = 0.95$) that sounds like a lot of fun . do you have any hobbies ?
A: C-AHED+RL	($r_{\text{target}} = 0.23$, $r = 0.84$) that is cool . i am a professional person .
A: REFERENCE	($r_{\text{target}} = 0.23$, $r = 1.0$) do you have any hobbies ?
A	that's great , i am a grad student
B	do you have any hobbies ? i really love sky diving , rollercoasters and cooking shows !
A	kayaking and reading psychology books
B	i'd go kayaking but i'm really not good at swimming so i'd be scared
A	i love it , i'm an awesome swimmer
B	do you like hummus ? it is a great alternative as a vegan , i love it !
A: AHED+CE	i do not have any pets , i do not have any pets
A: C-AHED+CE	($r_{\text{target}} = 0.28$, $r = 0.93$) i like to go to the beach , do you have any hobbies ?
A: C-AHED+RL	($r_{\text{target}} = 0.28$, $r = 0.95$) i like to go to the park and listen to music
A: REFERENCE	($r_{\text{target}} = 0.28$, $r = 1.0$) not really a fan , i am a meats kind of guy

VIII. CONCLUSION

We proposed a neural conversation model that can control the entrainment degree of generated responses according to the given entrainment degree. We applied reinforcement learning to optimize our model for automatic entrainment scores that incorporate local interpersonal distance as a reward. Experimental results showed the entrainment scores correlated with human assessments in both human-human and human-machine dialogues in a chit-chat domain. Our proposed model also generated comparable or more natural responses than conventional models based on the minimization of cross-entropy loss, while the degree of entrainment of the generated responses is well controlled.

Although our method outperformed the existing method based on cross-entropy loss, the entrainment degree of generated response can still be improved. This is because our method does not have any mechanism to explicitly access

the vocabulary used in the dialogue context on its decoding process. Hierarchical attention [43] or a copying mechanism [44] may explicitly solve this problem based on the word information in dialogue contexts. Incorporating such different aspects of entrainment as dialogue act choice is also important [8], [27].

A phenomenon where synchronization with each other in a dialogue, such as entrainment, is strongly related to the attributes of the dialogue participants and the relationships between them [45]. Therefore, we will focus on individuality or personality to control entrainment [21].

REFERENCES

[1] S. E. Brennan and H. H. Clark, "Conceptual pacts and lexical choice in conversation," *J. Experim. Psychol., Learn., Memory, Cognition*, vol. 22, no. 6, p. 1482, 1996.

[2] D. Reitter and J. D. Moore, "Predicting success in dialogue," in *Proc. 45th Annu. Meeting Assoc. Comput. Linguistics*, 2007, pp. 808–815.

[3] K. G. Niederhoffer and J. W. Pennebaker, "Linguistic style matching in social interaction," *J. Lang. Social Psychol.*, vol. 21, no. 4, pp. 337–360, Dec. 2002.

[4] M. Natale, "Convergence of mean vocal intensity in dyadic communication as a function of social desirability," *J. Personality Social Psychol.*, vol. 32, no. 5, p. 790, 1975.

[5] A. Ward and D. Litman, "Measuring convergence and priming in tutorial dialog," Univ. Pittsburgh, Pittsburgh, PA, USA, Tech. Rep., 2007.

[6] N. Campbell and S. Scherer, "Comparing measures of synchrony and alignment in dialogue speech timing with respect to turn-taking activity," in *Proc. 11th Annu. Conf. Int. Speech Commun. Assoc.*, 2010, pp. 1–4.

[7] Š. Beňuš, A. Gravano, R. Levitan, S. I. Levitan, L. Willson, and J. Hirschberg, "Entrainment, dominance and alliance in supreme court hearings," *Knowl.-Based Syst.*, vol. 71, pp. 3–14, Nov. 2014, doi: 10.1016/j.knosys.2014.05.020.

[8] M. Mizukami, K. Yoshino, G. Neubig, D. Traum, and S. Nakamura, "Analyzing the effect of entrainment on dialogue acts," in *Proc. 17th Annu. Meeting Special Interest Group Discourse Dialogue*, Sep. 2016, pp. 310–318.

[9] A. Nenkova, A. Gravano, and J. Hirschberg, "High frequency word entrainment in spoken dialogue," in *Proc. 46th Annu. Meeting Assoc. Comput. Linguistics Hum. Lang. Technol. Short Papers (HLT)*, 2008, pp. 169–172.

[10] R. Levitan, S. Benus, A. Gravano, and J. Hirschberg, "Entrainment and turn-taking in human-human dialogue," in *Proc. AAAI Spring Symp.*, 2015, pp. 44–51. [Online]. Available: https://www.aaai.org/ocs/index.php/SSS/SSS15/paper/view/10226

[11] M. Nasir, S. N. Chakravarthula, B. R. W. Baucom, D. C. Atkins, P. Georgiou, and S. Narayanan, "Modeling interpersonal linguistic coordination in conversations using word Mover's distance," in *Proc. Interspeech*, Sep. 2019, pp. 1423–1427, doi: 10.21437/Interspeech.2019-1900.

[12] B. Weiss, "Talker quality in interactive scenarios," in *Talker Quality in Human and Machine Interaction*. Cham, Switzerland: Springer, 2020, pp. 67–106.

[13] O. Vinyals and Q. Le, "A neural conversational model," 2015, *arXiv:1506.05869*. [Online]. Available: http://arxiv.org/abs/1506.05869

[14] I. V. Serban, A. Sordoni, Y. Bengio, A. C. Courville, and J. Pineau, "Building end-to-end dialogue systems using generative hierarchical neural network models," in *Proc. AAAI*, 2016, pp. 3776–3784.

[15] R. J. Williams, "Simple statistical gradient-following algorithms for connectionist reinforcement learning," *Mach. Learn.*, vol. 8, nos. 3–4, pp. 229–256, May 1992.

[16] A. Fandrianto and M. Eskenazi, "Prosodic entrainment in an information-driven dialog system," in *Proc. 13th Annu. Conf. Int. Speech Commun. Assoc.*, 2012, pp. 1–4.

[17] R. Levitan, "Entrainment in spoken dialogue systems: Adopting, predicting and influencing user behavior," in *Proc. NAACL HLT Student Res. Workshop*, 2013, pp. 84–90.

[18] Z. Tian, R. Yan, L. Mou, Y. Song, Y. Feng, and D. Zhao, "How to make context more useful? An empirical study on context-aware neural conversational models," in *Proc. 55th Annu. Meeting Assoc. Comput. Linguistics*, vol. 2, 2017, pp. 231–236.

[19] M. Ranzato, S. Chopra, M. Auli, and W. Zaremba, "Sequence level training with recurrent neural networks," 2015, *arXiv:1511.06732*. [Online]. Available: http://arxiv.org/abs/1511.06732

[20] M. Kusner, Y. Sun, N. Kolkin, and K. Weinberger, "From word embeddings to document distances," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 957–966.

[21] J. Li, M. Galley, C. Brockett, G. P. Spithourakis, J. Gao, and W. B. Dolan, "A persona-based neural conversation model," *CoRR*, vol. abs/1603.06155, pp. 1–10, Jun. 2016.

[22] C. Huang, O. Zaiane, A. Trabelsi, and N. Dziri, "Automatic dialogue generation with expressed emotions," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol.*, vol. 2, 2018, pp. 49–54. [Online]. Available: https://www.aclweb.org/anthology/N18-2008

[23] Y. Peng, Y. Fang, Z. Xie, and G. Zhou, "Topic-enhanced emotional conversation generation with attention mechanism," *Knowl.-Based Syst.*, vol. 163, pp. 429–437, Jan. 2019.

[24] G. Zhou, Y. Fang, Y. Peng, and J. Lu, "Neural conversation generation with auxiliary emotional supervised models," *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, vol. 19, no. 2, pp. 1–17, 2019.

[25] M.-T. Luong, H. Pham, and C. D. Manning, "Effective approaches to attention-based neural machine translation," 2015, *arXiv:1508.04025*. [Online]. Available: http://arxiv.org/abs/1508.04025

[26] C. Sankar, S. Subramanian, C. Pal, S. Chandar, and Y. Bengio, "Do neural dialog systems use the conversation history effectively? An empirical study," in *Proc. 57th Annu. Meeting Assoc. Comput. Linguistics*, 2019, pp. 32–37.

[27] S. Kawano, K. Yoshino, and S. Nakamura, "Neural conversation model controllable by given dialogue act based on adversarial learning and label-aware objective," in *Proc. 12th Int. Conf. Natural Lang. Gener.*, Oct./Nov. 2019, pp. 198–207.

[28] J. Li, W. Monroe, T. Shi, S. Jean, A. Ritter, and D. Jurafsky, "Adversarial learning for neural dialogue generation," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2017, pp. 2157–2169.

[29] L. Yu, W. Zhang, J. Wang, and Y. Yu, "Seqgan: Sequence generative adversarial nets with policy gradient," in *Proc. AAAI*, 2017, pp. 2852–2858.

[30] C. B. Browne, E. Powley, D. Whitehouse, S. M. Lucas, P. I. Cowling, P. Rohlfshagen, S. Tavener, D. Perez, S. Samothrakis, and S. Colton, "A survey of Monte Carlo tree search methods," *IEEE Trans. Comput. Intell. AI Games*, vol. 4, no. 1, pp. 1–43, Mar. 2012.

[31] M. J. Pickering and S. Garrod, "Alignment as the basis for successful communication," *Res. Lang. Comput.*, vol. 4, nos. 2–3, pp. 203–228, Oct. 2006.

[32] Y. Rubner, C. Tomasi, and L. J. Guibas, "A metric for distributions with applications to image databases," in *Proc. 6th Int. Conf. Comput. Vis.*, 1998, pp. 59–66.

[33] O. Pele and M. Werman, "A linear time histogram metric for improved sift matching," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2008, pp. 495–508.

[34] O. Pele and M. Werman, "Fast and robust Earth Mover's distances," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep. 2009, pp. 460–467.

[35] T. Arimoto, H. Sugiyama, M. Mizukami, H. Narimatsu, and R. Higashinaka, "Analysis of conversation topic between first-met speakers over multiple text-chats," *SIG-SLUD*, vol. 5, no. 3, pp. 66–71, 2019.

[36] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder–decoder for statistical machine translation," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2014, pp. 1724–1734.

[37] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980

[38] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, "BLEU: A method for automatic evaluation of machine translation," in *Proc. 40th Annu. Meeting Assoc. Comput. Linguistics (ACL)*, 2001, pp. 311–318.

[39] Y. Zhang, M. Galley, J. Gao, Z. Gan, X. Li, C. Brockett, and B. Dolan, "Generating informative and diverse conversational responses via adversarial information maximization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 1810–1820.

[40] C.-W. Liu, R. Lowe, I. Serban, M. Noseworthy, L. Charlin, and J. Pineau, "How NOT to evaluate your dialogue system: An empirical study of unsupervised evaluation metrics for dialogue response generation," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2016, pp. 2122–2132.

[41] C. Xing, W. Wu, Y. Wu, J. Liu, Y. Huang, M. Zhou, and W.-Y. Ma, "Topic aware neural response generation," in *Proc. 31st AAAI Conf. Artif. Intell.*, 2017, pp. 1–7.

[42] U. Olsson, F. Drasgow, and N. J. Dorans, "The polyserial correlation coefficient," *Psychometrika*, vol. 47, no. 3, pp. 337–347, Sep. 1982.

[43] C. Xing, Y. Wu, W. Wu, Y. Huang, and M. Zhou, "Hierarchical recurrent attention network for response generation," in *Proc. 32nd AAAI Conf. Artif. Intell.*, 2018, pp. 1–10.

[44] J. Gu, Z. Lu, H. Li, and V. O. K. Li, "Incorporating copying mechanism in sequence-to-sequence learning," in *Proc. 54th Annu. Meeting Assoc. Comput. Linguistics*, vol. 1, 2016, pp. 1631–1640.

[45] H. Giles, A. Mulac, J. J. Bradac, and P. Johnson, "Speech accommodation theory: The first decade and beyond," *Ann. Int. Commun. Assoc.*, vol. 10, no. 1, pp. 13–48, Jan. 1987.

SEIYA KAWANO received the B.E. degree from the National Institution for Academic Degrees and Quality Enhancement of Higher Education, in 2016, and the M.E. degree from the Nara Institute of Science and Technology (NAIST), Japan, in 2018, where he is currently pursuing the Ph.D. degree. His research interest includes natural language processing, especially dialogue systems. He is also a Research Fellow (DC2) of the Japan Society for the Promotion of Science.

MASAHIRO MIZUKAMI received the B.A. degree from Doshisha University, in 2012, and the M.S. and Ph.D. degrees in engineering from the Nara Institute of Science and Technology (NAIST), in 2014 and 2017, respectively. He is currently a Researcher with the NTT Communication Science Laboratories. His research interest includes spoken and natural language processing, especially on spoken dialogue systems.

KOICHIRO YOSHINO (Member, IEEE) received the B.A. degree from Keio University, in 2009, and the M.S. and Ph.D. degrees in informatics from Kyoto University, in 2011 and 2014, respectively. From 2014 to 2015, he was a Research Fellow (PD) of the Japan Society for the Promotion of Science. From 2015 to 2016, he was a Research Assistant Professor with the Graduate School of Information Science, Nara Institute of Science and Technology (NAIST). He is currently an Assistant Professor with NAIST. He is also a Researcher of PRESTO and JST, concurrently. His research interest includes spoken and natural language processing, especially on spoken dialogue systems. He is a member of ISCA, IPSJ, and ANLP. He received the JSAI SIG-research Award, in 2013.

SATOSHI NAKAMURA (Fellow, IEEE) received the B.S. degree from the Kyoto Institute of Technology, in 1981, and the Ph.D. degree from Kyoto University, in 1992. He was an Associate Professor with the Graduate School of Information Science, Nara Institute of Science and Technology, Japan, from 1994 to 2000. He was the Director of the ATR Spoken Language Communication Research Laboratories, from 2000 to 2008, and the Vice President of ATR, from 2007 to 2008. He was the Director-General of the Keihanna Research Laboratories and the Executive Director of the Knowledge Creating Communication Research Center, National Institute of Information and Communications Technology, Japan, from 2009 to 2010. He is currently the Director of the Data Science Center and a Professor with the Graduate School of Science and Technology, Nara Institute of Science and Technology, the Team Leader of the AIP Tourism Information Analytics Team, Riken, and a Honorary Professor with the Karlsruhe Institute of Technology, Germany. He is also the Director of the Augmented Human Communication Laboratory and a Full Professor with the Nara Institute of Science and Technology. He is one of the leaders of speech-to-speech translation research and has been serving various speech-to-speech translation research projects in the world, including C-STAR, IWSLT, and A-STAR. His research interests include the modeling and systems of speech-to-speech translation and speech recognition. He is a Fellow of ATR, IPSJ, and ISCA. He was elected as the Board Member of the International Speech Communication Association (ISCA), from 2011 to 2019, an Editorial Board Member of *IEEE Signal Processing Magazine*, from April 2012 to April 2015, and the IEEE SPS Speech and Language Technical Committee Member, from 2013 to 2016. He received the Yamashita Research Award, the Kiyasu Award from the Information Processing Society of Japan, the Telecom System Award, the AAMT Nagao Award, the Docomo Mobile Science Award, in 2007, the ASJ Award for Distinguished Achievements in Acoustics, and the LREC Antonio Zampolli Award 2012. He also received the Commendation for Science and Technology by the Minister of Education, Science and Technology and the Minister of Internal Affairs and Communications.

. . .