Heallo: Conversational System for Communication Training in Healthcare Professional Education

Xiang Zhang
Department of Computing
The Hong Kong Polytechnic University
Hong Kong, China
doris.x.zhang@connect.polyu.hk

Bruce X.B. Yu
Department of Computing
The Hong Kong Polytechnic University
Hong Kong, China
bruce.xb.yu@connect.polyu.hk

Yan Liu
Department of Computing
The Hong Kong Polytechnic University
Hong Kong, China
csyliu@comp.polyu.edu.hk

George Wing-Yiu Ng
Multi-disciplinary Simulation and Skills Centre
Queen Elizabeth Hospital
Hong Kong, China
georgeng@ha.org.hk

Nam-Hung Chia
Multi-disciplinary Simulation and Skills Centre
Queen Elizabeth Hospital
Hong Kong, China
cnhz01@ha.org.hk

Eric Hang-Kwong So
Multi-disciplinary Simulation and Skills Centre
Queen Elizabeth Hospital
Hong Kong, China
sohke@ha.org.hk

Sze-Sze So
Multi-disciplinary Simulation and Skills Centre
Queen Elizabeth Hospital
Hong Kong, China
sss083@ha.org.hk

Victor Kai-Lam Cheung
Multi-disciplinary Simulation and Skills Centre
Queen Elizabeth Hospital
Hong Kong, China
ckl414@ha.org.hk

2022 10th International Conference on Information and Education Technology (ICIET) | 978-1-6654-8048-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICIET55102.2022.9778989

***Abstract*—Effective communication is critical in the healthcare setting to ensure the safety and quality of patient care. To increase the accuracy and efficiency of communication, specific communication protocols have been developed. However, the current teaching of communication protocols remains in the theoretical realm, lack of effective training methods and practice environment. The purpose of this research is to assist communication training in healthcare professional education. We begin by designing communication tasks and collecting conversations on real-world handover cases. Then we propose Heallo – a computer conversational system that provides simulated scenarios for healthcare communication tasks. Specifically, Heallo converses with the user in a simulated communication task, assuming the role of the receiver. After the conversation, it will analyze the entire communication history and generate a detailed evaluation report based on the communication protocol. Heallo has received positive responses from healthcare professionals and has been incorporated into a hospital training program.**

Keywords—conversational system, communication training, healthcare education, natural language processing

I. INTRODUCTION

Communication in the healthcare setting, such as transferring information about a patient's state and care plan, is essential for safe and high-quality patient care [1]. Especially in team provision of healthcare, communication affects all aspects of care delivery, from the professional performance of skills to patient outcomes [2], and any missteps could result in severe consequences, such as delayed treatment, medication errors, and even mortality [3]. To help healthcare personnel communicate accurately and succinctly in different tasks, specific communication protocols have been developed, such as I-PASS [4] and SBAR [5].

Clinical handover is a typical task that requires effective communication between members of the healthcare team [6]. It is also a routine task that occurs numerous times throughout the day [7]. One of the most commonly used communication protocols in clinical handover is ISBAR (Identify-Situation-Background-Assessment-Recommendation), a standardized communication protocol recommended by the World Health Organization [8]. ISBAR has been shown of great potential to improve the transparency and accuracy of inter-professional and non-face-to-face handover in hospitals [9]. This protocol provides a systematic approach to handover by breaking the conversation down into five components: Identify, Situation, Background, Assessment, and Recommendation. Specifically, "Identify" usually appears at the beginning of a communication when you need to identify yourself, the patient, and the receiver; Following "Identify", the term "Situation" indicates the issue or reason for contact; In "Background", you can briefly summarize the patient's previous history relevant to the current problem; "Assessment" refers to the most recent clinical assessment, investigation, and your interpretation of the current situation; "Recommendation" is typically the concluding section, in which you can ask for advice or intervention and state your expectations. Following these steps, information can be delivered in a simplified and anticipated way between healthcare personnel.

However, the teaching of communication protocols remains in the theoretical realm, and there is scant evidence that theoretical abilities are transferred to practice [10]. ISBAR has not been well applied in real-world healthcare handover scenarios due to the lack of effective training. According to our interviews with healthcare professionals, this phenomenon is caused by several factors: First, conducting the exercise in a real healthcare setting is prohibitively expensive, as it requires patient consent and supervision by professionals; and experienced doctors are usually too preoccupied to practice communication with junior staff. Second, while theoretic results are straightforward to verify, practice processes are more difficult to monitor and evaluate. Practices between medical students generally lacks fidelity, and they usually can not get timely and accurate feedback. Although healthcare personnel are taught communication protocols in the classroom, without sufficient practice, they may be unable to deliver qualified communication, especially in emergent situations.

Enabling lifelike practice with low labor cost, conversational systems have been successfully used in automating training processes such as social skills training [11] and hotline counselor training [12]. Meanwhile, Natural Language Processing (NLP) techniques have led to transformative advances in smart healthcare [13], [14], which demonstrates the potential for healthcare communication training using NLP techniques.

Therefore, we are motivated to design a conversational system for low-cost and effective communication training in healthcare education. Specifically, in a simulated healthcare scenario (e.g., transferring the state and information of whom in an emergency situation), the user will be provided with all pertinent documents and practice communication via the conversational system. The system simulates the communication task by acting as the receiver and conversing with trainees. Following the conversation, a detailed evaluation report will be generated, indicating the communication's quality and areas for improvement.

Towards this goal, we develop the conversational system Heallo and use it for healthcare education. It has achieved satisfactory results among healthcare professionals and has also been successfully implemented in the Queen Elizabeth Hospital's summer intern training program.

II. Literature Review

Conversational systems have been studied for long with the first chatbot ELIZA [15] released in 1960s. Numerous conversational system applications have emerged recently in a variety of fields, including education, catering, and finance [16]–[18]. During the last half century, NLP technologies in developing conversational systems have varied from pattern matching, rule-based models, retrieve-based models, to generative models [19]. With the success of deep learning, neural networks start to be widely used for this task [20].

For deep learning methods, text data is first represented using word embedding, which projects sparse word representations into low-dimensional, dense vector representations [21]. Typical word embedding algorithms include word2vec [22], GloVe [23], and FastText [24]. Specifically, GloVe (Global Vectors for word representation) [23] word ebedding is trained on a huge corpus based on aggregated global co-occurrence statistics between words. Thus, the distance between two word vectors somehow reflects their semantic similarity.

After word embedding, different neural networks can be applied for downstream tasks, such as convolutional neural networks [25], recurrent neural networks [26], and Transformers [27]. BERT (Bidirectional Encoder Representations from Transformers) [28], one of the most popular pretrained models based on Transformers, has achieved state-of-the-art performance in many NLP tasks. The BERT-base model is constructed by 12 layers of transformer blocks with 100 million parameters, and pretrained over 3.3 billion word corpus [28]. It then can be used for a wide variety of NLP tasks simply by fine-tuning on the specific task without architecture modifications.

III. Methodology

We begin by developing a communication task based on real-world scenarios. In collaboration with healthcare professionals at Queen Elizabeth Hospital, we record conversations during these scenarios. Then conversational training system is designed using the collected data and the ISBAR protocol. The remainder of this chapter details the data collection process and system design.

A. Data Collection

The communication task is designed as the clinical handover in the event of an emergency, during which the doctor on duty is required to communicate the patient's situation to the senior doctor via non-face-to-face communication. Two typical cases are selected for the clinical handover task. The first one is a medical (MED) case concerning a patient with respiratory failure who may need elective intubation. The second one is a surgical (SURG) case, in which the patient might require emergency surgery for an acute abdominal injury. We cooperated with healthcare professionals in Queen Elizabeth Hospital to simulate the handover process on these cases. During the handover process, the duty doctor reported the case to a senior doctor based on the related documents including medical records, notes, and various testing reports (e.g., hematology reports, chemical pathology reports, CT scans). To protect patients' privacy, we fabricated all personal information in the conversations. Totally, we gathered 100 samples of conversation.

To perform annotation and processing, audio recordings were manually transcribed into text and segmented into sentences. Each sentence was labeled by three clinical experts based on the ISBAR protocol. The final dataset contained 1895 sentences from the duty doctor's side. Table I shows the distribution of sentences in each intent category.

TABLE I. Sentence Distribution in Each Intent Category

Intent	Number of sentences
Identify	230
Situation	97
Background	513
Assessment	762
Recommendation	293

Additionally, we asked healthcare professionals to list the critical components that are required to complete the task for each scenario. Each component can be associated with one or more keywords (each keyword contains a maximum of four words). Table II provides examples of critical components in the medical scenario. We have defined 26 critical components for the medical scenario and 29 for the surgical scenario in total.

TABLE II. Component Examples in the Medical Scenario

No.	Components	Keywords
1	Patient's name	Mr. Wong Hong Kin

No.	Components	Keywords
2	Bed number	Bed 24
3	Patient's age	69 years old
4	Admission time	14th February, yesterday
5	Respiratory failure	Respiratory failure, Increase breath
6	Ischemic heart disease	Ischemic heart disease, IHD
7	Blood test	ECG, ABG, CBC, CRP
…	…	…

B. System Design

The system's overview diagram is depicted in Fig. 1. It is divided into two stages: the conversation stage and the evaluation stage. In the conversation stage, the system assumes the role of a senior doctor and engages in the simulated communication task with healthcare personnel. Following the conversation, the entire communication history will be analyzed and graded during the evaluation stage. Technically, there are four components in the system: intent detector, information extractor, response retriever, and report generator.

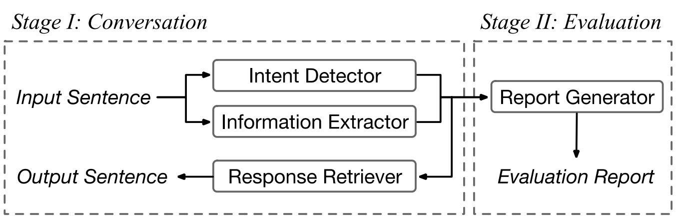

Fig. 1. The conversational system diagram.

1) Intent Detector: The intent detector assigns the input sentence to one of the ISBAR categories. We implement it by fine-tuning the BERT-base model [28] and connecting the output representation to a fully connected layer.

Given an input sentence, the BERT Tokenizer first segments it into a sequence of tokens. After that, we append the token [CLS] to the beginning of the sequence and pad it to a specified length. Then all tokens are converted into numbers by the predefined vocabulary of BERT-base. To determine the intent, the padded sequence is passed through the BERT-base model, resulting in a vector representation for each token. Then we connect the representation $\boldsymbol{r}$ of token [CLS], which can be regarded of as the representation of the entire sentence, to a fully connected layer and a SoftMax layer. So the prediction result $\boldsymbol{p}$ can be written as:

$$\boldsymbol{p} = \text{SoftMax}\,(\boldsymbol{r}W^T + \boldsymbol{b}) \tag{1}$$

where W and $\boldsymbol{b}$ are the parameters of the fully connected layer, and the predicted intent can be obtained by $y = argmax\,(\boldsymbol{p})$.

2) Information Extractor: The information extractor recognizes previously defined critical components from the input sentence. For each component, we will have a list of keywords as shown in Table II. We denote keyword k as a sequence of word embeddings $k = (\boldsymbol{v}_1, \dots, \boldsymbol{v}_t, \dots, \boldsymbol{v}_T)$, where T is the number of words in k, and $\boldsymbol{v}_t \in \mathbb{R}^{50}$ is a 50-dimensional word embedding of the t-th word. Here, we use glove.6B.50d [23] as the embedding dictionary, which is trained on Wikipedia 2014 and Gigaword5 with 6 billion tokens and 400 thousand vocabularies. Then we can obtain a semantic embedding for the keyword k by averaging over word embeddings:

$$\boldsymbol{e}_k = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{v}_t \tag{2}$$

For an input sentence, we first generate the corresponding 1-gram, 2-gram, 3-gram, and 4-gram sequences. For example, the 4-gram sequences of sentence "*He requires 10 ml dopamine.*" would be ("*He requires 10 ml*", "*requires 10 ml dopamine*", "*10 ml dopamine.*"). All of the words in these four sequences are combined to create a set of n-gram words. Following the same procedure of keyword embedding and Equation 2, we can obtain the semantic embedding $\boldsymbol{e}_w$ of the n-gram word w. The similarity of keyword k and n-gram word w is then denoted by the cosine similarity of their semantic embeddings:

$$s(k, w) = \cos\,(\boldsymbol{e}_k, \boldsymbol{e}_w) = \frac{\boldsymbol{e}_k \boldsymbol{e}_w}{\|\boldsymbol{e}_k\|\|\boldsymbol{e}_w\|} \tag{3}$$

In this way, we can determine the keyword that is most similar to the n-gram word w. And if the similarity is greater than the threshold, w will be marked as the corresponding critical component of the keyword. After processing all the words in the n-gram set, the information extractor returns words associated with a particular component and their components as extracted information.

3) Response Retriever: When a user enters a sentence, the response retriever selects an appropriate response from the response pool and outputs it. We have defined responses for a few specific intents and components. Once they are detected in the input sentence, the corresponding response will be retrieved. And for the remaining intents and components, a set of generic responses is prepared, from which the retrieving program will randomly select one.

4) Report Generator: The report generator evaluates the whole conversation based on the results from the intent detector and the information extractor, and generates an evaluation report for the user. According to the consultation of healthcare experts at Queen Elizabeth Hospital and the simulated grading of the collected conversation samples, we finally formulate three criteria:

- *Category Number (CN)* indicates the number of intent categories covered during the conversation (a maximum of 5 categories in the ISBAR protocol). For instance, if a conversation contains categories I, B, A, and R, then the *CN* value is equal to four.

- *Wrong Order (WO)* indicates how many intents are in the wrong order, namely not following the I-S-B-A-R sequence. Similar to the idea of edit distance, *WO* is calculated as the minimum number of intents deleted from a sequence to keep the correct order. For example, the *WO* of intent sequence I-S-B-A-B-R-A-R is 2.

- *Missed Information (MI)* indicates the number of critical components that have not been mentioned in the conversation.

Based on the above three criteria, Table III shows the categorical grading scheme.

TABLE III. CATEGORICAL GRADING SCHEME FOR COMMUNICATION EVALUATION

CN	WO×5+MI	Grade
5	0 ~ 5	Excellent
5	6 ~ 15	Very Good
5	> 15	Good
4	0 ~ 20	
4	> 20	Fair
3	0 ~ 15	
3	>15	Poor
< 3	≥ 0	

IV. RESULTS AND DISCUSSION

In this section, we demonstrate the web application of Heallo by a use case and discuss its usage in healthcare professional education.

A. Data Collection

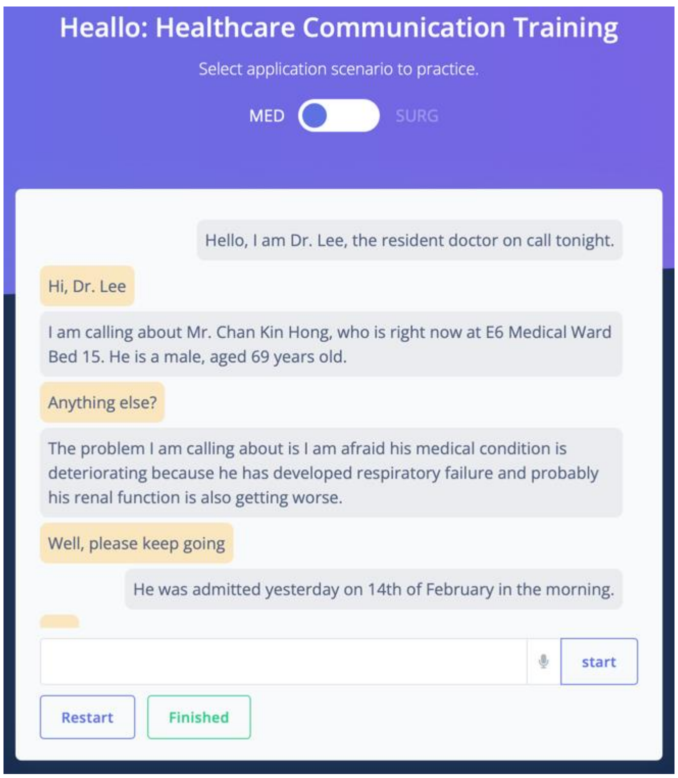

Fig. 2. Conversation interface of Heallo.

Fig. 2 shows the conversation interface of Heallo. Above the chatting box is a button for selecting the practice scenario, which can be medical or surgical. The gray boxes on the right contain user inputs, while the yellow boxes contain system outputs. As can be seen from the dialogue, the outputs are based on the results of the intent detection and the information extracted from the previous sentence. Typically, the system's output sentence contains little information and serves primarily as a guide for the conversation to continue. When the conversation is complete, the user can proceed to the evaluation stage by clicking the green "Finished" button at the bottom.

B. Evaluation

In the evaluation stage, Heallo will generate an evaluation report (see Fig. 3) based on the intent of each sentence and information extracted during the conversation. The evaluation report begins by assigning an overall grade to the conversation in accordance with the rules introduced in Section III-B4. The conversation is then summarized from two perspectives: Part I details the intents detected during the dialogue process and whether the order of the intents follows the ISBAR protocol; Part II identifies the critical components missing from the conversation. Based on this feedback, users can quickly identify their communication weaknesses, thereby achieving the training objective.

Fig. 3. Evaluation report generated by Heallo.

C. Usage in Education

We conducted three rounds of testing with senior doctors at Queen Elizabeth Hospital during the development of Heallo, and the final version achieved satisfactory results. It was then implemented as part of the summer intern simulation training at Queen Elizabeth Hospital. According to the feedback forms returned by a total of 40 participants, the average rating for satisfaction with training methods is 4.38 (5 for very satisfied and 1 for very dissatisfied). And the overall satisfaction with the program's effectiveness is 4.65. In the free-form feedback, many respondents expressed a desire for additional simulation scenarios. Based on these findings, we believe Heallo has provided effective clinical communication training on the ISBAR protocol. In the future, we will continue to improve the system in response to educational requirements and user feedback.

V. CONCLUSION

In this work, we propose a practical solution for communication training in healthcare education. In collaboration with healthcare professionals, we carefully designed communication tasks in medical and surgical scenarios. Conversations in both scenarios are collected and annotated based on the ISBAR protocol. Then we create Heallo, a conversational system that simulates environment for the healthcare communication task. It is capable of conversing as the receiver in a communication task and providing detailed feedback afterward. Heallo has received positive feedback from both healthcare experts and junior doctors. We hope that this initial attempt of applying NLP technologies in the healthcare communication training will promote the development of communication training systems and inspire researchers in both NLP and healthcare education.

ACKNOWLEDGMENT

This work was supported by Innovation and Technology Fund (ITS/110/19) from the Innovation and Technology Commission of Hong Kong. We would also like to express our gratitude to all healthcare professionals who participated in this program for their active involvement and helpful advice.

REFERENCES

[1] H. Ratna, "The importance of effective communication in healthcare practice," *Harvard Public Health Review*, vol. 23, pp. 1–6, 2019.

[2] C. Foronda, B. MacWilliams, and E. McArthur, "Interprofessional communication in healthcare: An integrative review," *Nurse education in practice*, vol. 19, pp. 36–40, 2016.

[3] M. Leonard, S. Graham, and D. Bonacum, "The human factor: the critical importance of effective teamwork and communication in providing safe care," *BMJ Quality & Safety*, vol. 13, no. suppl 1, pp. i85–i90, 2004.

[4] A.J.Starmer,N.D.Spector,R.Srivastava,A.D.Allen,C.P.Landrigan, T. C. Sectish, I.-P. study group, *et al.*, "I-pass, a mnemonic to standardize verbal handoffs," *Pediatrics*, vol. 129, no. 2, pp. 201–204, 2012.

[5] K.M.Haig,S.Sutton,andJ.Whittington,"Sbar:asharedmentalmodel for improving communication between clinicians," *The joint commission journal on quality and patient safety*, vol. 32, no. 3, pp. 167–175, 2006.

[6] M.Botti,T.Bucknall,P.Cameron,M.-J.Johnstone,B.Redley,S.Evans, and S. Jeffcott, "Examining communication and team performance during clinical handover in a complex environment: the private sector post-anaesthetic care unit," *Medical Journal of Australia*, vol. 190, no. S11, pp. S157–S160, 2009.

[7] M. Broekhuis and C. Veldkamp, "The usefulness and feasibility of a reflexivity method to improve clinical handover," *Journal of evaluation in clinical practice*, vol. 13, no. 1, pp. 109–115, 2007.

[8] W. P. Safety and W. H. Organization, "Patient safety curriculum guide: Multi-professional edition," 2011.

[9] S. Marshall, J. Harrison, and B. Flanagan, "The teaching of a structured tool improves the clarity and content of interprofessional clinical communication," *BMJ Quality & Safety*, vol. 18, no. 2, pp. 137–140, 2009.

[10] S. Buckley, L. Ambrose, E. Anderson, J. J. Coleman, M. Hensman, C. Hirsch, J. Hodson, D. Morley, S. Pittaway, and J. Stewart, "Tools for structured team communication in pre-registration health professions education: a best evidence medical education (beme) review: Beme guide no. 41," *Medical teacher*, vol. 38, no. 10, pp. 966–980, 2016.

[11] H. Tanaka, H. Negoro, H. Iwasaka, and S. Nakamura, "Embodied conversational agents for multimodal automated social skills training in people with autism spectrum disorders," *PloS one*, vol. 12, no. 8, p. e0182151, 2017.

[12] L. K. Fryer, K. Nakao, and A. Thompson, "Chatbot learning partners: Connecting learning experiences, interest and competence," *Computers in Human Behavior*, vol. 93, pp. 279–289, 2019.

[13] S.Velupillai,H.Suominen,M.Liakata,A.Roberts,A.D.Shah,K.Morley, D. Osborn, J. Hayes, R. Stewart, J. Downs, *et al.*, "Using clinical natural language processing for health outcomes research: overview and actionable suggestions for future advances," *Journal of biomedical informatics*, vol. 88, pp. 11–19, 2018.

[14] B. Zhou, G. Yang, Z. Shi, and S. Ma, "Natural language processing for smart healthcare," *arXiv preprint arXiv:2110.15803*, 2021.

[15] J. Weizenbaum, "Eliza—a computer program for the study of natural language communication between man and machine," *Communications of the ACM*, vol. 9, no. 1, pp. 36–45, 1966.

[16] R. Dale, "The return of the chatbots," *Natural Language Engineering*, vol. 22, no. 5, pp. 811–817, 2016.

[17] S. Chandel, Y. Yuying, G. Yujie, A. Razaque, and G. Yang, "Chatbot: efficient and utility-based platform," in *Science and Information Conference*, pp. 109–122, Springer, 2018.

[18] R. M. Thomas, S. Punna, M. C. K. K. Reddy, and B. RamanaMurthy, "Survey on artificially intelligent chatbot," *Journal of Applied Science and Computations*, vol. 6, no. 1, pp. 85–94, 2019.

[19] R.Yan,Y.Song,X.Zhou,andH.Wu,"Shallibeyourchatcompanion?: Towards an online human-computer conversation system," in *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, pp. 649–658, ACM, 2016.

[20] R. Yan, "" chitty-chitty-chat bot": Deep learning for conversational ai.," in *IJCAI*, vol. 18, pp. 5520–5526, 2018.

[21] T. Mikolov, É. Grave, P. Bojanowski, C. Puhrsch, and A. Joulin, "Advances in pre-training distributed word representations," in *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, 2018.

[22] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Advances in neural information processing systems*, pp. 3111– 3119, 2013.

[23] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pp. 1532– 1543, 2014.

[24] A.Joulin,E.Grave,P.Bojanowski,M.Douze,H.Jégou,andT.Mikolov, "Fasttext. zip: Compressing text classification models," *arXiv preprint arXiv:1612.03651*, 2016.

[25] Y.ZhangandB.C.Wallace,"Asensitivityanalysisof(andpractitioners' guide to) convolutional neural networks for sentence classification," in *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pp. 253–263, 2017.

[26] A. Bhargava, A. Celikyilmaz, D. Hakkani-Tür, and R. Sarikaya, "Easy contextual intent prediction and slot detection," in *2013 ieee international conference on acoustics, speech and signal processing*, pp. 8337– 8341, IEEE, 2013.

[27] A.Vaswani,N.Shazeer,N.Parmar,J.Uszkoreit,L.Jones,A.N.Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *NIPS*, 2017.

[28] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 4171–4186, 2019.